RISK ASSESSMENT OF FEDERAL FEMALE OFFENDERS

A Thesis Submitted to the College of
Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy
In the Department of Psychology
University of Saskatchewan
Saskatoon

By

CATHERINE A. STEWART

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Psychology
University of Saskatchewan
Saskatoon, Saskatchewan, S7K 5A5

ABSTRACT

Much remains to be learned about predicting recidivism in women, including which risk assessment instruments may be most beneficial and what other factors may add to a more complete understanding of recidivism for women. The predictive validities of four risk assessment instruments (i.e., the Level of Service Inventory- Revised (LSI-R), the Level of Service and Case Management Inventory (LS/CMI), the Violence Risk Scale (VRS) and the Statistical Information on Recidivism – Revised 1 (SIR-R1) scale) were compared in an archival sample of 101 federal female offenders on a series of outcome measures including institutional misconduct and general and violent recidivism. The incremental predictive validity of a number of gender informed variables (e.g., history of abuse, economic concerns, childcare responsibility and self harm/suicidality) was also explored. The four risk assessment instruments performed well (AUCs from .650 to .842), adding additional evidence to the growing literature supporting the use of traditional risk assessment instruments with female offenders (Smith, Cullen, & Latessa, 2009; Yang, Wong, & Coid, 2010; Andrews, et al., 2011). Few of the gender informed variables correlated with outcome with the exception of childhood emotional abuse which was related to violent recidivism (AUC .687). However, consistent with the pathways and gender-responsive theories of criminal behavior (Daly, 1992; Reisig, Holtfreter, & Morash, 2006), the combination of childhood emotional abuse and substance abuse appeared the most promising composite of gender informed variables as it was predictive of revocation, institutional misconduct and general and violent recidivism ($r = .208$ to $.248$, $p < .05$). Moreover, for most of the risk assessment instruments their predictive utility was improved by adding a gender informed variable or composite, although which variable or composite in particular was not consistent across outcomes or risk assessment instruments. Implications and future directions for risk assessment of women are discussed.

ACKNOWLEDGMENTS

Firstly, I would like to thank my supervisor, Dr. Wormith for his support and guidance throughout my graduate student career and especially with this research. His experience, good humour, and patience were invaluable. I would also like to thank my committee members, Drs. Farthing, Woods, and Wong, and the external examiner, Dr. Brown, for their enthusiasm about this work. Their insightful questions, suggestions, and recommendations have enriched this research and final project immensely. Thanks also to the Correctional Service of Canada for allowing me access to a rich source of information from which to build my data and to Research Department at the Regional Psychiatric Centre (RPC), especially Dr. Deqiang Gu, for facilitating data collection and sharing their expertise throughout this endeavor. I also appreciate the Psychology Department at Edmonton Institution for finding me a temporary home when I could not be at the RPC. Thank you to Tessa Halldorson and Curtis Brad who were Research Assistants for this project and gave generously of their time. I would also like to express my gratitude to the Centre for Forensic Behavioural Sciences and Justice Studies for their financial support of this research.

Secondly, I would like to thank my extended family and friends who helped me reach this goal in too many ways to list, but without whom I know I would not have succeeded. Although often from afar, I always felt their support and encouragement. Special thanks to my brother, Dan, for being someone I can look up to (in more ways than one!) and for always being there for me. Also, warmest thanks to my grandparents who, although no longer with us, provided me with a strong foundation from which to grow and were always four of my biggest supporters.

Finally, deepest thanks to my parents, who embraced my curiosity from the beginning, valued higher education, taught me compassion and humour, and had faith in me even when my own wavered. I dedicate this dissertation to them. Character built. (Well, phase one anyway!)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APD: Antisocial Personality Disorder

CSC: Correctional Service Canada

LS/CMI: Level of Service/Case Management Inventory

LSI-OR: Level of Service Inventory: Ontario Revision (earlier version of the LS/CMI)

LSI-R: Level of Service Inventory- Revised

OMS: Offender Management System

PCL-R: Psychopathy Checklist Revised

PIC-R: Personal, Interpersonal, and Community-reinforcement Theory

RNR: The Risk, Need and Responsivity Principles of Effective Correctional Treatment

SIR-R1: Statistical Information on Recidivism - Revised

SIR/GSIR: General Statistical Information on Recidivism

VRAG: Violence Risk Appraisal Guide

VRS: Violence Risk Scale

INTRODUCTION

1.1 Risk Assessment in Federal Female Offenders

Risk assessment is fundamental to correctional psychology. It is used to assist in predicting the likelihood of reoffence, as well as, the needs of an offender when entering the correctional system. In addition, risk assessment may be used to guide security placements, and various kinds of community release. It may also be used to identify offenders' continuing needs upon release. However, given the lack of research on risk assessment in female federal offenders, no risk assessment instruments are currently used as common practice by Correctional Service of Canada (CSC) with these women (Blanchette, 2000). While communications within CSC have suggested that the use of the Level of Service Inventory- Revised (LSI-R; Andrews & Bonta, 1995) for federal female offenders is defensible based on its predictive utility in provincial and American samples (F. Bellemare, personal communication, September 9, 2003), without validation in federal Canadian samples, this practice is unlikely to become policy. The limited support for the use of risk assessment instruments with federal female offenders leaves correctional workers vulnerable to using their judgment alone to evaluate the risk these women represent, as well as the needs they require addressed.

The current study aimed to explore needs/risk assessment in federal female offenders. It has been known for decades that clinical intuition and professional judgments alone cannot be assumed to produce accurate assessments (Meehl, 1954). While much research has been conducted on actuarial risk assessment, it has primarily focused on male offenders (see Gendreau, Little & Goggin, 1996 for a review), and its generalizability to female offenders had been questioned (Blanchette, 2000; Hannah-Moffat & Shaw, 2001). Although it has been widely assumed that female and male offenders differ in a variety of important ways, there is a paucity of empirical research to support this contention (Giordano & Cernkovich, 1997) and increasing support for the use of traditional risk assessment instruments with female offenders in general (Smith, Cullen, & Latessa, 2009; Yang, Wong, & Coid, 2010; Andrews, et al., 2011). However, little research to date has specifically focused on the use of risk assessment instruments for the more serious federal female offender group (Folsom & Atkinson, 2007; Nafekh & Motiuk, 2002). Exploring whether actuarial risk assessment instruments, developed for and with male offenders, will have comparable success with federal female offenders, and if other factors

thought to relate to female recidivism will augment the predictive utility of these instruments could help increase our understanding of female recidivists.

1.2 Female Offenders

A "snapshot" of Canada's federal offender population (those serving a sentence of at least 2 years) demonstrated that the 379 females incarcerated at that time represented approximately 3% of the total inmate population (CSC, 2005). Eighty-three percent of these women were serving their first federal sentence, and 73% of women were serving a sentence for a violent crime. While this frequency of violence seems high, it is important to note that women receive only about 11% of the charges for violent offences that occur in Canada (Hatch & Faith, 1989). Further, the CSC snapshot showed that thirty-seven percent of women were serving a sentence of less than three years, while another 31% were serving a sentence of three to six years and 18% of these women were serving life.

In a review of the literature of characteristics of female offenders, the Federally Sentenced Women Program (1994) described the majority of female inmates as survivors of abuse and trauma as perpetrated by their families of origin or their intimate partners. Education and employment difficulties were also prominent concerns of female offenders. In fact, most of the female offenders were identified as poor, lacking marketable skills, and single parents solely responsible for childcare. They were described as dependent on men, welfare and alcohol, the latter of which was identified as being often comorbid with eating disorders, affective disorders, and possible post- traumatic stress disorder.

With regard to success in the community post release, Dowden and Andrews (1999) estimated the overall recidivism rates in their meta-analysis of female offender treatment effectiveness were 43% for treated offenders and 57% for offenders in the control condition. A lower recidivism rate of 22% was demonstrated in a sample of first-time Canadian federal female offenders by Belcourt, Nouwens and Lefebvre (1993), but this study focused on returning to federal custody and did not consider less serious reconvictions. By comparison, in a relatively large scale study of male (3267) and female (81) offenders, Bonta, Lipinski, and Martin (1993) found that 36% of women reoffended in 3 years, as compared to 49% of men. They suggested that a 3-year follow-up period would capture 90% of recidivists. Gobeil and Robeson Barrett (2007) found that between 37.2% and 38% of female offenders were revoked while on conditional release, but that only 4.2% to 5.2% of federally sentenced women were reconvicted

for a violent offence within two years. While there is much variability in recidivism rates for these women, reoffending is not uncommon by any estimate.

1.3 Personal, Interpersonal, and Community-reinforcement Theory

The risk, need and responsivity principles of effective correctional treatment (RNR) form the theoretical basis of most of the common risk assessment instruments and stem from Andrews and Bonta's (2003) Personal, Interpersonal, and Community-reinforcement Theory (PIC-R). These principles have also received support for their application to female offenders (see Dowden & Andrew, 1999, for a meta-analytic review). PIC-R is a predictive theory of criminal conduct that takes a broad social learning theory perspective with influences from behavioural, and self-control theories (Blanchette & Brown, 2006). PIC-R, drawing on behavioural theories, postulates that variation in human behaviour is explained by influences that are present in the immediate situation of action (Andrews & Bonta, 2003). These influences are the costs and rewards of behaviour and the antecedents of behaviour that signal the rewarding or costly consequences of actions. As with control theory, PIC-R builds on the human desire to seek pleasure and avoid pain (Andrews & Bonta, 2003). According to Andrews and Bonta (2003), antecedents included such things as: others as models of cost and reward, cognitions, affective states, physical resources, perceptions, and past experience. Variation in these influences lead to the acquisition, maintenance, and modification of human behaviour (Andrews & Bonta, 2003).

Situated in personality and social science, PIC-R recognizes that these influences of behaviour are developed, maintained and modified by personal, interpersonal, and community factors (Andrews & Bonta, 2003). Some factors are automatic (e.g., sensory changes resulting from drug use), while others are personally and interpersonally mediated (e.g., self-management and self reward, and external approval, respectively). These mediated influences can also be created and maintained by the larger social, economical and political system in which the individual lives (Andrews & Bonta, 2003). PIC-R states that all criminal acts are under behavioural control from four main sources: the individual (personally mediated), others in the situation (interpersonally mediated), the act itself (automatic factors), and other aspects of the situation (e.g., opportunity; Andrews & Bonta, 2003; Blanchette & Brown, 2006).

Bonta and Andrews (2003) suggested that long-term change is most likely and efficient when the personal, interpersonal and community influences at the root of the immediate influences of behaviour are targeted by treatment, not the immediate influences themselves. PIC-

R not only targets influences of criminal conduct, but also aims to reinforce and increase motivation for pro-social behaviours (Andrews & Bonta, 2003). As the rewards for prosocial behaviours increase, there is more for the individual to lose by committing criminal acts.

PIC-R is presented as a theory to understand criminal conduct generally, irrespective of gender, class or ethnicity. While gender is recognized as a factor influencing both the person and the situation, it is not central to the theory (Blanchette & Brown, 2006). However, its utility to predict and reduce recidivism has received support with female offenders given the applicability of the PIC-R-rooted risk, need, and responsivity principles (Dowden & Andrews, 1999).

The *risk principle* refers specifically to the amount of treatment or attention an offender should receive, such that those with more identified needs are at greater risk to reoffend and will benefit from more intensive treatment, while those who are low risk will not, and in fact can experience a negative treatment effect (Andrews, 2000; Andrews & Bonta, 2003). Risk assessment may include dynamic risk factors, which are factors that have the potential to change (e.g., attitudes), and static factors (e.g., age) that cannot be targets of change (Gendreau, et al., 1996). Dowden and Andrews' (1999) meta-analytic review found stronger treatment effects for higher risk samples, than for lower-risk samples. More specifically, for studies with predominantly female participants, recidivism was reduced by 19% in the higher risk group, while no treatment effect was found in the lower risk group. This effect increased to a 24% reduction in recidivism for the higher risk groups when studies that targeted females exclusively were examined. However, Dowden and Andrews (1999) were not able to monitor the intensity of the treatment provided, and instead compared only treatment vs. no treatment. This method only partly addresses the risk principle, namely that treatment can reduce recidivism for higher risk offenders, but does not for lower risk offenders. Their study did not look at a gradient of treatment in response to increasing risk levels, whereby low risk offenders receive less treatment than high risk offenders. Finally, Dowden and Andrews classified females into higher and lower risk groups, based on current and past involvement with the correctional system, not a complete actuarial assessment per se. Blanchette and Brown (2006) argue that the higher risk groups may have been much more likely to demonstrate a reduction in recidivism than the lower risk groups given their higher base-rate of offending prior to treatment. In fact, they point out that some lower risk groups were comprised of nonoffenders, highlighting the need for additional exploration of the applicability of the risk principle with female offenders.

An offender's needs can be divided into criminogenic needs and noncriminogenic needs (Andrews, 2000). Criminogenic needs include factors, both attributes of the offender and his or her situation, which have been found to directly increase one's likelihood to reoffend (e.g., anger control difficulties). The *need principle* refers to targeting interventions at an offender's identified criminogenic needs in order to reduce recidivism (Andrews, 2000). Noncriminogenic needs include factors that are not directly related to one's risk for reoffence (e.g., Mental health care; Andrews, 2000; Blanchette & Brown, 2006), but may be problematic for the individual. These noncriminogenic factors may also require intervention to ensure the well-being of the offender, but as sole treatment goals, they will not lead to a reduction in recidivism (Andrews, 2000).

Once again, Dowden and Andrews (1999) meta-analytic review of treatment effectiveness for female offenders, supports the needs principle. That is, in studies of predominantly female offenders, treatment programs that targeted criminogenic needs versus noncriminogenic needs significantly reduced recidivism by 26%. However, when looking at specific criminogenic needs, only two of four criminogenic factors significantly reduced recidivism. The strongest predictors of recidivism were interpersonal factors, like family process and antisocial associates, and personal factors, like antisocial cognitions and skills deficits, with reductions in recidivism ranging from 30 to 38%. School or work focused interventions, as well as substance abuse interventions did not significantly reduce recidivism (despite receiving support as criminogenic needs for male offenders; see Gendreau, Little, & Goggin, 1996 for a meta-analytic review), and were in fact slightly negatively correlated with reductions in recidivism.

As Blanchette and Brown (2006) pointed out, the need principle *itself* is not generally questioned in the literature, but instead questions arise about which factors are criminogenic for women. Andrews and Bonta (2003) described the "big four" risk and need factors as history of antisocial behavior, antisocial personality, antisocial cognitions, and antisocial associates which have been shown to be reliably the strongest predictors of recidivism for a variety of offender populations. Substance abuse, the quality of family/martial relationships, education/employment opportunities, and leisure activities have repeatedly been demonstrated to have moderate correlations with recidivism and complete what they refer to as the "central eight" (Andrews & Bonta, 2003). There is evidence that criminogenic factors that are important for male offenders

are also important for female offenders (Simourd & Andrews, 1994; Dowden & Andrews, 1999; Blanchette & Brown, 2006; Andrews, et al., 2011). However, most recently in a study including seven samples of female offenders, Andrews and colleagues (2011) have suggested including substance abuse as one of the "big five" for women (in lieu of the "big four" for men) given the salience of that factor for this population. Others have also suggested gender differences in criminogenic needs more broadly (McClellan et al., 1997; Dowden & Andrews, 1999; Funk, 1999; Heilbrun, et al., 2008; Van Voorhis, Wright, Salisbury, & Bauman, 2010), and thus further investigation may be warranted, especially in the less studied population of federal female offenders. A more in depth review of potential criminogenic risk/need factors for female offenders follows.

The third principle of effective correctional treatment stemming from PIC-R is the *responsivity principle*, which refers to treatment that is both capable of effecting change on the targeted needs and is appropriately matched to offenders learning styles (Andrew, Zinger, Hoge, Bonta, Gendreau & Cullen, 1990). The responsivity principle is further divided into general and specific responsivity. *General responsivity principle* suggests that treatment outcome will be best with service delivery that includes structured, behavioural interventions (Andrews et al., 1990). The effective treatments for male offenders are generally based on behavioural and social learning theories and aim to develop skill and change cognitions through modeling, rehearsal and reinforcement (Andrews et al., 1990), which have also been found to be effective service delivery methods for female offenders (Dowden & Andrews, 1999). The *specific responsivity principle* focuses on case-specific factors that may help or hinder treatment effectiveness (Andrews et al., 1990). These factors can be internal (e.g., cognitive ability, motivation) or external (e.g., therapist characteristics; Kennedy, 2001). Gender is conceptualized as a specific responsivity factor (Andrews & Bonta, 2003).

1.4 Pathways or Gender-Responsive Models for Female Offenders

The pathways or gender-responsive theories of criminal behaviour are supported by feminist scholars and stem initially from Daly's (1992) work which classified female offenders into five pathways or trajectories explaining the life events that contributed to them coming into conflict with the law. These pathways included the "street women" pathway where girls left an abusive home, engaged in petty crime or prostitution, and eventually turned to drug use or addiction in response to life on the street which in turn led to more criminal involvement. The

"harmed and harming women" pathway was where girls were abused or neglected and began to "act out", then later became violent when abusing alcohol or in response to psychological difficulties or trouble coping effectively. These women may also have had difficulties with drug addiction. The "battered women" pathway included women who were involved in or recently ended abusive relationships. The "drug connected women" pathway included women who used or sold drugs through a boyfriend or family member. The "other" pathway included women with immediate economic concerns or greed who did not fit into the other four described pathways. Reisig, Holtfreter, and Morash (2006) later argued that this "other" pathway may have represented female offenders who were economically motivated and thus were more similar to their criminally oriented male counterparts.

There has been much contention in the literature between the RNR and the pathways or gender-responsive theories (see Hannah-Moffat, 2009; Morash, 2009; Smith, Cullen, & Latessa, 2009; and Taylor & Blanchette, 2009 for a recent review of this debate). The pathways or gender-responsive proponents have argued that mainstream criminological theories, like RNR, are androcentric and have ignored important elements unique to female criminality, like abuse histories, substance abuse, mental health issues, and economic marginalization (Daly, 1992; Benda, 2005; Reisig, Holtfreter, & Morash, 2006). Even if some of these elements overlap with RNR criminogenic needs, they are purported to be experienced differently by women (Hannah-Moffat, 2009; Morash, 1999). Further, they have highlighted differences in offending patterns (Hannah-Moffat, 2009), such as women committing fewer crimes and being less violent, and voiced concerns about over and under-classifying women offenders with RNR derived assessment tools, especially those women identified as having gendered pathways to crime (Reisig, Holtfreter, & Morash, 2006). In addition to the differences between these theories of crime for women, Morash (2009) also described that pathways theories are not concerned with prediction, as is RNR, but instead with describing the ways that gender affects people's lives and behaviours. Further, Morash (2009) explained that unlike RNR, the needs addressed by treatment consistent with a gender-responsive theoretical orientation would not be confined to those related to recidivism and would be targeted at women who are and are not high risk offenders. Both of these differences have significant policy and practice implications.

Hannah-Moffat (2009) also noted that these two theories are being integrated into correctional policy and programming with increasing regularity, despite their fundamental

differences. Hubbard and Matthews (2008) attempted to reconcile some to these differences in a recent review article on female young offenders. They suggested that gender-responsive knowledge about the experience of gender helps clarify the RNR responsivity principle as it applies to girls. However, the focus on gender as a responsivity issue exclusively has also been criticized as responsivity issues are seen as having lower priority than criminogenic needs (Hannah-Moffat, 2009). Further, this criticism underscores that the intended goals of these theories in correctional treatment (i.e., reductions in recidivism versus improved quality of life) need to be reconciled if the two are to be combined successfully. Similar challenges have been faced in reconciling the Good Lives Model and RNR (see Andrews, Bonta, & Wormith, 2011, who called for the prioritizing of reductions in recidivism within the correctional system).

Some efforts have also been made to combine knowledge from these RNR and pathways or gender-responsive theories to reduce female reoffending. Specifically, some researchers have begun looking at supplementing RNR oriented risk/needs assessments with gender-responsive variables. While researchers have called for the construction of a gender-informed assessment tool that is created and validated with women from its inception (Blanchette & Brown, 2006; Hannah-Moffat, 2009), such supplements have been supported as an appropriate interim measure (Taylor & Blanchette, 2009). Wright, Salisbury, and Van Voorhis (2007) found that a gender-responsive scale (including measures of childhood abuse, low relationship support, depression/anxiety and psychosis) performed as well as a gender-neutral scale (with constructs like those in risk assessment instruments derived from RNR) when added to a static risk scale to predict institutional misconduct. In addition, the combination of all three scales had the most predictive power. Further, the addition of gender-responsive factors to static risk items was also found to improve the prediction of reoffending in community samples (Van Voorhis, Salisbury, Wright, & Bauman, 2008). Moreover, the gender responsive factors related to institutional misconduct were not the same as those for reoffending (Van Voorhis et al., 2008). Finally, while adding gender-responsive factors to the LSI-R improved the prediction of serious institutional misconduct, it did not better predict community outcomes (Salisbury, Van Voorhis, & Spiropoulos, 2009) until the gender-responsive factors were selected for specific offender samples (Van Voorhis, Wright, Salisbury, & Bauman, 2010). Ongoing refinement, validation, and replication will be required to create a supplement or supplements that can be generalized to other female offender samples, but efforts to date have been promising.

1.5 Female Offenders and Predicting Recidivism

For decades, it has been a struggle to classify female offenders into groups to better understand and address their common needs and risk, while still maintaining respect for the heterogeneity of the individual women within each group (Widom, 1978). Women in correctional settings have long been viewed differently from their male counterparts. Women's aggression is often characterized as expressive, while men's is seen as an instrumental attempt to gain control (Shaw & Dubois, 1995). Primarily, women are thought to behave criminally in response to having been victimized throughout their lives (Shaw & Dubois, 1995), while men's behaviour is not likewise explained.

While research on female offenders has offered insight about contributors to the development of female criminality (e.g., among the most common, are childhood and current family situation, peer relationships, psychopathology, and schooling and economic variables; see Bloom, Owen & Covington, 2003; Loeber & Stouthamer-Loeber, 1998; Giordano & Cernkovich, 1997; Caspi, Lynam, Moffitt, & Silva, 1993) only select variables have been examined in the context of recidivism as being of particular importance for female offenders. These gender-informed variables are described below.

1.5.1 Gender-informed criminogenic risk/need variables.

It is important to note that although gender-informed criminogenic risk/need variables are being treated as discrete from the variables included in some of the more common male-derived actuarial risk assessment instruments, this distinction is an over simplified dichotomy. Substance abuse, for example, will be reviewed in both sections, while economic/financial concerns are explored with the risk/need variables from the current risk assessment instruments. Similarly, self-harm and suicidality may also overlap with emotional control and mental health concerns.

1.5.1.1 History of victimization.

The prevalence of traumatic and stressful life experiences is high for offenders, with familial abuse and adult victimization being higher for female than male offenders (Carlson & Shafer, 2010). Research on juvenile offenders has shown child abuse to be significant not only in the development of criminality, but more specifically in female (but not male) reoffending (Funk, 1999). While studies have supported a relationship between early victimization (childhood and/or lifelong abuse) and criminal behaviour in women, the relationship is viewed to

be very complex (Correctional Service of Canada, 1994; McClellan, Farabee & Crouch, 1997; Rettinger, 1998; Benda, 2005; Salisbury & Van Voorhis, 2009). Rettinger (1998) found that physical, sexual, and emotional abuse as a child, and physical abuse as an adult were each predictive of general recidivism, while sexual and emotional abuse in childhood and physical abuse in adulthood also predicted violent recidivism. However, these relationships were no longer significant after controlling for general risk/need (Rettinger & Andrews, 2010; Lowenkamp, Holsinger, & Latessa, 2001). It is also important to note that 81% of women who were sexually and emotionally abused as children did not recidivate violently, nor did 83.1% of women physically abused as adults (Rettinger, 1998). Loucks and Zamble (2000) found only preadolescent sexual abuse to be predictive of violent reoffending, but not the overall frequency of reoffence, and suggested that this reoffending may be context specific (e.g., domestic violence).

Women with histories of victimization may also be at increased risk for other factors related to reoffending. For example, McCartan & Gunnison (2010) found that female offenders who were sexually abused were more likely to be in conflict with the law as juveniles, had more difficulty staying employed, had criminal peers, and were more likely to be in an abusive relationship than offenders who did not experience such abuse. A longitudinal study of low-income minority participants also found that childhood maltreatment was related to adult arrest conviction, but was largely mediated by high school graduation for girls (Topitzes, Mersky, & Reynolds, 2011).

Moreover, substance abuse (a well accepted criminogenic need for male offenders, see Gendreau, Little, Goggin, 1996, for a meta-analytic review) was found to be a better predictor of criminal behaviour in women than early victimizations and was also found to be related to early victimizations (McClellan, et al., 1997). Thus, it was postulated that early victimization experiences led to an increased risk for substance abuse that then increased the risk of criminal behaviour in these women (McClellan, et al., 1997; Rettinger, 1998). However, McCartan & Gunnison (2010) did not find substance abuse was more common with female offenders with a sexual abuse history relative to those without. Further, Bonta, Pang and Wallace-Capretta (1995) found substance abuse was not predictive of recidivism in a sample of federally incarcerated women, while, surprisingly, adult abuse victimization correlated negatively with recidivism. However, Widom, Schuck, & White (2006) found that problematic alcohol use mediated the

relationship between childhood victimization and violent arrests for women. Taken together, these findings suggest that further exploration of abuse variables in present family and family of origin, and their relationship with substance abuse and risk for recidivism is warranted.

1.5.1.2 Self-injury and suicidality.

Bonta and colleagues (1995) also found that a history of self-injury was an important predictor of general recidivism for female offenders, while Rettinger & Andrews (2010) found self-abuse to be a significant predictor of violent recidivism. Blanchette (1997a) agreed that self-injury behaviours deserved further attention, given the disproportionate representation of mental illness in female offenders, and that nearly 50% of female offenders have attempted suicide (Loucks & Zamble, 1994). Women offenders who attempt or complete suicide are also more likely to have a history of self-harm than those who show no suicidal behaviours (Serin, Motiuk & Wichmann, 2002). Even if such concerns are not directly related to future reoffending, they require clinical attention (Loucks & Zamble, 2000) and their treatment may facilitate further recidivism prevention programming.

1.5.1.3 Childcare responsibility.

Women in general carry more responsibility for childcare than men, and have less training and education, and generally work in lower income positions (Bloom et al., 2003). The socio-economic challenges faced by women may be more likely to result in further recidivism when familial demands are present, suggesting a further potential criminogenic need to be explored. Brown and Motiuk (2005) found that limited parenting skills predicted readmission to custody for all federal offenders, but more so for women than men. Bonta and colleagues (1995) also found that childcare responsibilities were predictive of recidivism, but only when women were single parents. And, while Rettinger (1998) found that parenting was not predictive of recidivism in provincial female offenders, this may have been due to the fact that she combined being a single parent, living with one's children and having childcare concerns in her analysis of parenting. Gover, Perez, and Jennings (2008) also did not find a significant relationship between having a child and institutional misconduct, but suggested focusing on whether women had their children in their care prior to incarceration to improve the precision of this measure.

1.5.2 Risk/criminogenic need domains from current risk assessment instruments.

While the following domains have received significant support in the research literature on male offenders (Andrews & Bonta, 2003; Gendreau, Little, & Goggin, 1996; Andrews &

Bonta, 1995; Andrews, Bonta, & Wormith, 2004; Cormier, 1997; Wong & Gordon, 1998-2003), the empirical support for female offenders is not as extensive.

1.5.2.1 Criminal and social history.

Generally, criminal and social history predictors represent static variables. That is, once achieved they represent increased risk for reoffence and cannot be addressed through treatment. The influence of criminal and social history have been defined in numerous ways, focusing on current and previous offences (e.g., type, number, pattern), age at various points of contact with the criminal justice system, security status (e.g., escape history, institutional misconduct, security classification), life situation (e.g., marital/employment status upon admission, number of dependents) and stability of upbringing. Moreover, these historic variables have been explored in relation to various reoffence outcomes (e.g., general recidivism, violence recidivism, reconviction, rearrest, etc.). The following is a discussion of some of the findings that pertain to female offending.

In general, criminal history had been found to be predictive of recidivism for a number of female offender samples (Andrews et al., 2011) as has violent recidivism in a meta-analytic study by Collins (2010). Female offenders tend to be older than male offenders when they first come into contact with law (Kratcoski & Scheijerman, 1974), and at the time of their first violent incident (de Vogel & de Ruiter, 2005). Age at first conviction and admission to a penitentiary have been shown to be inversely related to general recidivism in federal female offenders (Bonta et al.,1995). Further, violent female recidivists have been shown to be significantly younger at first conviction than nonrecidivists, and at first violent offence than other first time female offenders (Rettinger, 1998, Weizmann-Henelius, Viemero, & Eronen, 2004). However, a younger age at the time of offence was not predictive of recidivism for female drunk drivers (Lapham, Skipper, Hunt, & Chang, 2000), but this is a specific and very small subset of female offenders.

In general, a previous history of criminal behavior differentiated between recidivists and nonrecidivists in samples of provincial female offenders (Andrews et al., 2011). However, in a sample of federally sentenced women, Bonta and colleagues (1995) found that a history of juvenile delinquency was not predictive of recidivism, but prior adult convictions and prior incarceration were positively associated with recidivism.

More specifically, examinations of the types of previous offences committed by female offenders have yielded some interesting predictive information. Bonta and colleagues (1995) found that federal female offenders with previous drug-related charges were less likely to be reconvicted, while property crimes had no relationship to recidivism and those convicted of robbery offences were more likely to be reconvicted. However, unlike for male offenders, offence type did not relate to institutional misconduct for female offenders (Gover, Perez, & Jennings, 2008). Of the studies of female offenders conducted to date, none has evaluated a history of violent sexual offending, and Bonta and colleagues (1995) have explained that too few women have such a history to evaluate.

As more serious or violent offences are generally associated with longer sentences, while property offences often earn shorter sentences, it is not surprising that sentence length has been shown to positively predict violent recidivism (Collins, 2010). However, it was negatively related to general recidivism and at the extreme, Bonta and colleagues (1995) also found that serving a sentence of life imprisonment was also associated with reduced recidivism.

While other areas of social history overlap with dynamic criminogenic needs described below (e.g., stability of intimate relationships, educational attainment), stability of family upbringing is a risk area that remains static. Brown and Motiuk (2005) found that limited childhood ties, negative maternal relations as a child, difficulties in parental relationships and family criminality were related to readmission in a sample of federal female offenders. Similarly, Dowden and Brown (2002) suggested that parental substance abuse history may be a proxy for negative family background, and it was found to be predictive of recidivism (based on 12 effect sizes) in their meta-analytic review of substance abuse programming for federal female offenders.

1.5.2.2 Education/employment.

Blanchette (2002) described the education and employment status of federally sentenced women, confirming the applicability of this need area to this population. She reported that 48% of women in federal facilities had less than 10 grade education as compared to 19% of the general adult population of Canada, while 80% of women were unemployed when admitted to a federal institution, as compared to only 10% of adults in the general population. In addition, 24.7% of federal female offenders had employment identified as a considerable need upon admission, and 48.9% were identified as having some need for improvement (Blanchette,

1997a). Gover and colleagues (2008) also found that lower educational attainment was related to increased institutional misconduct for women. Further, Brown and Motiuk (2005) argued that while lower educational attainment predicted readmission to custody for all federal offenders, the relationship was stronger for women relative to their male counterparts. Violent provincial female recidivists also scored higher on education/employment needs than women who did not reoffend violently (Rettinger, 1998).

Completion of a General Equivalency Diploma was associated with longer survival times outside of prison for female offenders; however, completing vocational-technical training while incarcerated was associated with shorter survival times outside of prison in a sample of female offenders in Oklahoma (Brewster & Sharp, 2002). This may have had to do with the marketability of the skills learned or the offenders' interest in the vocational programs offered and underscores the importance of considering how interventions will carry through to the community. An unstable employment history was related to recidivism for female federal offenders, as was limited job skills and dissatisfaction with her trade or profession (Brown & Motiuk, 2005).

In contrast, while fewer British female recidivists had paid employment (12%) post-release than nonrecidivists (approx. 20%), employment was not significantly related to recidivism in this sample (Morris & Wilkinson, 1995). Moreover, Bonta and colleagues (1995) also found that employment variables did not differentiate recidivists from nonrecidivists in a sample of federal female offenders; however, having nonemployment sources of financial support, like an illegal income or welfare support did relate to higher risk for reoffending, highlighting the complexity of female offenders' criminogenic needs.

1.5.2.3 Family and intimate relationships.

Blanchette (1997a) described that 23.1% of federal female offenders had family/marital issues identified as a considerable need upon admission, while another 47.3% were identified as having some need for improvement in this area. Similarly, violent provincial recidivists scored higher on family/marital needs than women who did not reoffend violently (Rettinger, 1998).

Women identified partners and parents as their major providers of practical help and advice (Strauss & Falkin, 2001). They explained that their partners were "a sympathetic ear", while their parents affirmed their self-worth. Other family members also provide a significant amount of support to female offenders, making up about half of women's support networks

(Strauss & Falkin, 2001). Further, federal female offenders who acknowledged discord within their family unit or problematic family ties were at increased risk of readmission (Brown & Motiuk, 2005).

1.5.2.4 Companions/peers.

Seventeen percent of federal female offenders had associates identified as a considerable need upon admission, while 63.7% were identified as having some need for improvement in this area (Blanchette, 1997a). Violent provincial recidivists also scored higher on criminal companions needs than women who did not reoffend violently (Rettinger, 1998). Moreover, Brown and Motiuk (2005) found that federal female offenders who associated with substance abusers or had many criminal associates or friends were at increased risk of readmission within three years. However, Bonta and colleagues (1995) found that whether the offence occurred with an associate was not predictive of reoffence.

1.5.2.5 Substance abuse.

Thirty-eight and a half percent of female offenders had substance abuse identified as a considerable need upon admission to a federal institution, while 23.6% were identified as having some need for improvement (Blanchette, 1997a). In an evaluation of substance abuse programming for female offenders, Dowden and Brown (2002) found that substance abuse factors were moderately predictive of general recidivism. More specifically, combined drug and alcohol abuse, and drug abuse alone were the strongest predictors of recidivism, followed by alcohol abuse alone. Violent provincial recidivists scored higher on drug/alcohol problems than women who did not reoffend violently (Rettinger, 1998). Further, in addition to mediating the pathway from childhood victimization, as described above, Widom and colleagues (2006) also found a direct path between substance abuse (which appeared to result from early aggressive behavior) and violent arrest in women. Moreover, Brown and Motiuk (2005) found that drug abuse predicted readmission to custody for all federal offenders, but that the relationship was stronger for female offenders than males. Alcohol abuse was also linked to increased risk of readmission for these women. In contrast, Bonta and colleagues (1995) found that alcohol and drug abuse were not predictive of general recidivism for federal female offenders. Unexpectedly, substance abuse charges were negatively associated with recidivism, although mildly, in Dowden and Brown's (2002) study, and the authors suggested these offenders may not actually have had a substance abuse problem despite their offence. Finally, Dowden and Blanchette (2002) found

substance abuse treatment significantly reduced general recidivism by 18%, but not violent recidivism, possibly due to the low base rate of violent reoffence.

1.5.2.6 Emotional/personal.

Personal/emotional needs (including impulsivity, emotional control deficits, mental health concerns etc.) were identified as a considerable need upon admission for 35.7% of federal female offenders, while 52.2% of federal female offenders were identified as having some need for improvement (Blanchette, 1997a). Most women sentenced to prison experienced multiple life events characterized by loss in the 12 months preceding their incarceration, and many, if not most, endorsed feelings of tension, anxiety and worry (Keaveny & Zauszniewski, 1999).

Female offenders, when compared to a matched sample of male offenders, were more often diagnosed with Borderline Personality Disorder, and less with Antisocial and Narcissistic Personality Disorders (de Vogel & de Ruiter, 2005). However, Luberto, Zavatti, and Gualandri (1997) stated that patients affected by a mental illness present a lower risk of deviance compared to the general population, and often commit crimes that are less serious. Yet, repeat violent offenders were more than twice as likely as first time offenders to have Borderline Personality Disorder diagnosis (Weizmann-Henelius, Viemero, & Eronen, 2004). Repeat violent female offenders were also more likely than one-time violent or nonviolent offenders to report greater acting out and less inhibition when angry (Verona & Carbonell, 2000). Finally, in a large sample of federal offenders, limited coping abilities were more strongly related to readmission for women than men (Brown & Motiuk, 2005). They also found federal female offenders were at increased risk of readmission when they had difficulties with solving interpersonal problems, impulsivity, callousness, frustration tolerance, risk taking, thrill seeking, and manipulation (Brown & Motiuk, 2005).

1.5.2.7 Criminal attitudes/cognitions.

Blanchette (1997a) described that 8.8% of federal female offenders had attitudes identified as a considerable need upon admission, and 20.9% of female offenders were identified as having some need for improvement in this area. By comparison, attitudes represent a considerable criminogenic need for 39% of male offenders, and are also considered one of the most valuable treatment targets for male offenders as they are highly correlated to recidivism (Blanchette, 2002). Higher criminal attitudes/orientation needs differentiated recidivists from nonrecidivists for both general and violent recidivism in a sample of provincial offenders

(Rettinger, 1998). Moreover, in a study of juvenile female offenders, nonviolent offenders were more delinquent in orientation, and endorsed more rule breaking and nonconforming on the Jesness Inventory than violent offenders (Allen, Rupert, Spatafora, Windell, Gaulier, and Conti, 2002).

More specifically, criminogenic attitudes often refer to offenders' views and thoughts about their crimes, sentences, and supervision. Henning, Jones, and Holdford (2005) found that female domestic violence offenders were more likely than males domestic violence offenders to place greater blame on their partners than themselves for the offence. They also found that a significant number of women denied the incident or minimized the severity of the offence in some way. In addition, substance abuse treatment following driving while intoxicated showed significant attitudinal changes in female offenders' willingness to accept responsibility for substance use and driving behaviour (Juhnke & Sullivan, 1995). With regard to sentencing, the vast majority of female offenders felt their sentences were too harsh, and 53% were dissatisfied with lawyers' services in Kratcoski and Scheijerman study of American offenders (1974). With regard to supervision, Sandu, Dodder, and Davis (1990) found significantly more women than men had an unfavourable attitude toward community supervision. Finally, federal female offenders were more likely to be readmitted to custody if they had negative attitudes toward the law, police, corrections, or community supervision (Brown & Motiuk, 2005). Having values that were generally non-conforming, disrespectful of property or supportive of violence were also associated with increased readmission for these women (Brown & Motiuk, 2005).

1.5.2.8 Community functioning.

Community functioning includes leisure, accommodation, finance, support, deportment (e.g., hygiene), and health needs (Blanchette, 2002). Blanchette (1997a) described that 10.4% of federal female offenders had community functioning identified as a considerable need upon admission, while 58.8% were identified as having some need for improvement in this area. Moreover, federal female offenders appeared to be more likely to be readmitted to custody when lacking leisure activities than their male counterparts (Brown & Motiuk, 2005).

Accommodation inadequacy was related to readmission in a sample of federal female offenders (Brown & Motiuk, 2005). Accommodation was also found to be predictive of general recidivism in provincial female offenders (Rettinger, 1998). Moreover, violent provincial recidivists scored higher on accommodation needs than women who did not reoffend violently

(Rettinger, 1998). Finally, in a sample of British female offenders, 50% of recidivists reported unsatisfactory accommodations, while only 14% of nonrecidivists reported likewise (Morris & Wilkinson, 1995).

Similarly, in the same British sample, 50% of recidivists reported financial management difficulties compared to only 25% of nonrecidivists, and almost one third of the recidivists reported that drug use was a drain on their financial resources (compared to only 3% of nonrecidivists; Morris & Wilkinson, 1995). The same was true for alcohol, with almost 25% of recidivists reporting it as a financial drain, as compared to 4% or nonrecidivists. However, Rettinger (1998) found that having social assistance as a main source of income was unrelated to recidivism for provincial female offenders, while for federal female offenders it has been linked to readmission (Brown & Motiuk, 2005). Moreover, Holtfreter, Reisig, & Morash (2004) found that after accounting for poverty, LSI-R was not a significant predictor of rearrest for female offenders. Thus, they stressed the impact of poverty and suggested the LSI-R does not capture the economic marginality of female offenders sufficiently, and although there were a number of methodological difficulties with their study (e.g., high attrition rate, self-reported outcome), economic pressures appear to warrant further investigation.

1.5.2.9 Criminal personality/antisocial pattern.

Criminal personality and antisocial patterns often refer to persistent criminal lifestyles, and are conceptually similar to the construct of psychopathy, often coinciding with factor 1 (e.g., callousness, lack of remorse, pathological lying, manipulativeness, etc.) and factor 2 (e.g., boredom, impulsivity, criminal versatility, etc.), respectively. Antisocial Personality Disorder (APD) has also been explored as a measurement of criminal personality. The Psychopathy Checklist Revised (PCL-R) is the most common instrument for assessing psychopathy in offenders and will be reviewed later with other male-derived risk assessment instruments. However, here we will discuss both criminal personality and patterns of antisocial behaviour as criminogenic needs.

It is important to understand the criminal personality may appear differently in female offenders than male offenders. For example, female offender scores on the PCL-R are lower on average than the scores of male offenders (Weizmann-Henelius, Viemero, & Eronen, 2004). As well studies have shown a lower base rate of psychopathy in women than men, as measured by the PCL-R (9% in Vitale, Smith, Brinkley, and Newman, 2002, and 12.9% in Salekin, Rogers,

Ustad, & Sewell, 1998). It has been suggested that this may be due to the prevalence of psychopathy in the population being truly lower for women, as it is seen to be with other antisocial behaviour based disorders like APD and conduct disorder, or may be due to the PCL-R items not adequately capturing the construct in females (Vitale, Smith, Brinkley, and Newman, 2002).

Despite these difference, violent female recidivists are significantly more likely to have APD, and significantly higher PCL-R scores (both factors) than first time offenders (Weizmann-Henelius, Viemero, & Eronen, 2004). Both APD and psychopathy have deception and social norm violation in common (Warren & South, 2006), while shallow affect, lack of remorse, conning/manipulative and lack of realistic goals, callous/lack of empathy, impulsivity, and failure to accept responsibility for own actions seem to be important indicators of psychopathy in female offenders (Weizmann-Henelius, Viemero, & Eronen, 2004). In incarcerated female offenders, APD has been found to be associated with impulsivity, aggressive and irresponsible behaviour, higher rates of childhood abuse, and is more likely to be comorbid with other personality disorders, while psychopathy has been associated with higher rates of property crimes, previous incarcerations, and a lack of remorse (Warren & South, 2006).

Only PCL-R factor 1, reflecting personality characteristics (e.g., pathological lying, manipulativeness, lack of remorse, etc.), significantly predicted recidivism in a sample of female offenders (r=0.26; Salekin et al., 1998). Richards and colleagues (2003) found factor 1 was also related to decreased time in community before reoffence, as was the total PCL-R score. Factor 2, although not found to be predictive of recidivism for female offenders, is moderately predictive for male psychopaths (Salekin et al.,1998), and violent provincial female recidivists scored higher on antisocial pattern subscale of the Level of Service Inventory-Ontario Revision (LSI-OR) than women who did not reoffend violently (Rettinger, 1998).

1.5.2.10 Patterns of violent behaviour.

Patterns of violent behaviour, including violent lifestyle, violence while institutionalization, weapon use, and violence cycles, have been used to predict future violence in men, and while some of the above criminogenic needs have been assessed on violent female recidivists, patterns of violence have rarely, if ever, been explored. Of these behaviours, institutional misconduct has probably received the most attention. Repeat violent female

offenders were more likely than one-time violent or nonviolent offenders to show more aggression while incarcerated (Verona & Carbonell, 2000).

One study by Koons-Witt and Schram (2003) report that female offenders were more likely to be involved with incidents of violence when personal weapons (i.e., hands, feet, teeth) and knives are used, as opposed to blunt objects or guns. However, Bonta and colleagues (1995) found that weapon use was not predictive of recidivism for federal female offenders.

1.5.2.11 Release and reintegration.

Security classifications may (partly) reflect the risk/needs presented by an offender in the institution, and security classification at release has been explored as a proxy for risk for reoffence at release. However, women rated on the Security Reclassification Scale for Women and the Offender Security Level did not significantly differ on rates of return to custody or reoffence, whether they had received minimum, medium, or maximum security classifications, (Blanchette, 2005) and it is argued that institutional interventions may have mitigated the between group differences.

Release type has also been explored and release on full parole was predictive of reduced recidivism (Bonta and colleagues, 1995), which may reflect the characteristics of offenders who qualify for full parole, as opposed to day parole, statutory release or release at warrant expiry, or possibly the benefit of the reintegration assistance provided by full parole. However, it is important to note that release decisions are complex, can impact outcome, and do not reflect a client attribute like most risk/need factors. Finally, women report less self-efficacy than men to remain abstinent from substances in high-risk situations (Pelissier & Jones, 2006), further underscoring the importance of considering release environment conditions when preparing women for reintegration.

1.6 Current Male Derived Risk Assessment Instruments

Although female-specific theories of recidivism will provide promising avenues of exploration for risk assessment, the current risk assessment instruments developed on male offenders may also provide additional insight into general risk/needs that are common to both genders (Loucks & Zamble, 2000). While claiming that such instruments are gender-neutral and applying them to female offenders without validation would be a mistake, their utility can be empirically determined. The following is a review of the more common actuarial instruments and their application to female offenders.

A revision of the General Statistical Information on Recidivism scale (SIR-R scale; adapted from Nuffield, 1982) is currently being used by the National Parole Board of Canada to predict recidivism and is completed for each nonaboriginal, male offender in CSC's system. While the SIR-R scale is not to be used with female or aboriginal offenders (Correctional Service Canada, 2003) as sufficient research has yet to be conducted, given its widespread use with the remaining male population, it could be a useful addition to the current exploration. This scale focuses on static factors, primarily on the criminal history of the offender, but also includes variables such as marital status, number of dependents and employment at time of arrest. Bonta and colleagues (1995), rating the original General Statistical Information on Recidivism scale (GSIR) from file information, found that the scale did significantly predict recidivism overall in female offenders, albeit mildly, but it did so without a linear progression from one category to the next, such that increased ratings did not necessarily lead to increased recidivism. In addition, the sample of women examined was small (97), employment at time of arrest and number of dependants were not scored as data not available, and some of the items in the SIR scale that were not found to be predictive were concepts that literature has previously supported in relation to female risk assessment. These include substance abuse and employment (Blanchette, 1997a, b), suggesting that further validity studies are required before this instrument is deemed inappropriate. Nafekh & Motiuk (2002) found more promising results for the overall utility of the SIR-R1 with a larger sample of female offenders, as its total score was significantly related to readmission to federal custody within three years.

Other research on female offenders has focused on another well-established instrument to assess risk factors for recidivism among offenders, the LSI-R. The LSI-R examines an offender's level of criminogenic risk in a range of life and social areas such as employment, finances, companions, substance abuse etc. in order to give the assessor insight into an offender's current situation (Andrews & Bonta, 1995). Rettinger (1998) found that provincial female offenders had mean LSI-R scores similar to those typically reported for male offenders and that the LSI-R was a significant predictor of recidivism, correctly classifying 84% of nonrecidivists and 73% of recidivists. An earlier study by Coulson, Ilacqua, Nutbrown, Giulekas and Cudjoe (1996) found that female offenders had lower average LSI-R scores than male offenders, yet the LSI-R continued to be a significant predictor of recidivism.

In fact, a recent meta-analysis of the LSI-R, including 27 effect sizes for female offenders, confirmed the utility of this risk assessment instrument for women in general (Smith, Cullen, & Latessa, 2009). Holtfreter and Cupp's (2007) earlier review of the empirical evidence challenged the "gender neutrality" of the LSI-R. They highlighted lower effect sizes and mean scores for women relative to men, studies that controlled for gender instead of examining men and women separately, and the diversity of the female samples as cautions against generalizing these findings to women offenders more widely. However, Andrews and colleagues (2011a) in a review of meta-analytic summaries reported a weighted mean effect size for women (.345) that is at least as strong as that which they reported for men (.322). Further, they attribute Holtfreter and Cupp's (2007) modest validity to low investigator allegiance (which they suggested as a proxy for assessment integrity) and short follow-up periods.

The only study to date to examine federal female offenders (Folsom & Atkinson, 2007) found a self-report version of the LSI-R predicted general recidivism such that those outside of the low risk group (with a total score greater than 12.99 out of a possible 53) were significantly more likely to reoffend. Violent recidivism was not related to LSI-R scores, but this may have been due to the low incidence of violence in the study. This provides initial support for the utility of this risk assessment instrument with federal women offenders and, as mentioned above, its use was recently encouraged by CSC given its anticipated potential with this population (F. Bellemare, personal communication, September 9, 2003). Exploring clinician rated LSI-R scores is a logical next step.

Comparably, the LSI-OR (now the Level of Service/Case Management Inventory, LS/CMI, Andrews, Bonta & Wormith, 2004), a revision of the LSI-R, has also been shown to predict general and violent recidivism well with community and institutional male offenders (Girard & Wormith, 2004). Further, it correctly identified 82% of nonrecidivist and 74% of provincial female recidivists, and may also prove useful with federal women offenders (Rettinger, 1998). In fact, more recent studies rarely differentiate between the Level of Service instruments when discussing their utility (see Andrews and colleagues' (2011a) review of meta-analytic summaries as an example), given the substantial item overlap between these assessment instruments.

The predictive values of the LSI-R and the LSI-OR were not augmented when variables suggested in the literature by Shaw (1994) to be uniquely relevant to female offenders (e.g.

living arrangements and supports for the females in question, number of children/dependents and where they currently reside, education and employment aspirations, family background including history of abuse, and mental health concerns) were added. Interestingly, in a meta-analysis conducted by Dowden and Andrews (1999) treatment targets similar to the Shaw variables, in particular, family process variables (e.g. affection and supervision), were the strongest predictors of reduced reoffending for female offenders. It is possible that the LSI-R/OR captured these family process variables indirectly by examining whether parental and spousal relationships were rewarding. However, this contention should be explored further.

Among the many risk assessment instruments that have been developed for male offenders, a meta-analysis suggested that the LSI-R was the most useful actuarial instrument of its kind for predicting recidivism in male offenders (Gendreau, et al., 1996). It was later found to be the single best predictor of institutional misconduct (Gendreau, Goggin, & Law, 1997). A more recent study of the Psychopathy Checklist –Revised (PCL-R; Hare 1991) and the LSI-R confirmed that the LSI-R was a more comprehensive instrument and surpassed the predictive value of the PCL-R for general recidivism, yet when predicting violent recidivism the LSI-R was only moderately more effective (Gendreau, Goggin & Smith, 2002). Gendreau and colleagues (1996) explained that the LSI-R owes its success to the fact that it focuses on a group of dynamic factors that are thought to be more appropriate targets for treatment, such as antisocial cognitions and values. Further, their meta-analysis revealed that the PCL-R was the best instrument of its kind focusing on anti-social personality factors and is recommended for clinicians particularly interested in predicting violence. Moreover, they suggested that for special populations the PCL-R could be most valuable when used in addition to the LSI-R.

The PCL-R has shown promise for use with female offenders in assessing psychopathy and its relationship with recidivism and institutional misconduct despite observed gender differences (Salekin, Rogers, & Sewell, 1997; Salekin et al.,1998; Vitale & Newman, 2001; Richards, Casey, & Lucente, 2003). However, CSC does not support the use of this measure with female offenders, thus it was not practical to include in the current study of this population. Similarly, the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice & Cormier, 1998), another risk assessment instrument designed to predict violent recidivism, which also been shown to be predictive of institutional misconduct and nonviolent convictions with male offenders (Kroner & Mills, 2001), requires rating of the PCL-R in its scoring and thus was also

excluded from the present study. Moreover, a recent study by Hastings and colleagues (2011) found the VRAG was not predictive of institutional misconduct or general or violent recidivism for women (as it was for men), and suggested it was inadvisable to use the VRAG as it provided no more information than psychopathy measures alone and could in fact distort and decrease the accuracy of the assessment of risk for women.

A fifth instrument employed when assessing needs/risk is the Violence Risk Scale (VRS; Wong & Gordon, 1998-2003). The unique component of this instrument is that it includes an assessment of change, while determining an offender's risk for violent reoffence. Like the LSI-R, it focuses on static or historical and dynamic factors, but can capture change by assessing the dynamic risk factors both before and after treatment, resulting in a more complete assessment of the individual's potential to be influenced by effective treatments. To do this, readiness to change is evaluated pre-treatment and post-treatment and used to adjust post-treatment risk scores. While the present study does not have a post-treatment phase, the VRS can be used as a stand-alone measure of risk and given the importance placed on effective programming for female offenders, understanding the utility of this instrument for female offenders pre-treatment would be an important first step in its validation.

To date, VRS has been shown to be predictive of non-violent and violent recidivism for provincial and federal male offenders (Wong & Gordon, 2006). It was predictive of institutional violence with one sample of mentally disorder offenders in the United Kingdom (Dolan & Fullam, 2007) and with the same sample with the addition of 11 women (Dolan, Fullam, Logan, & Davies, 2008), but not consistently in another small UK forensic inpatient sample (Grevatt, Thomas-Peter, & Hughes, 2004). This discrepancy, where small effects were only found for the subscales of the VRS once the frequency of physical assaults (but not the prevalence) was examined, may be due in part to prorating at least four of the 20 dynamic risk items relating to community and release, and to the shorter follow up period of 6 months for misconduct. The VRS has also been used with violent-prone federal offenders (Wong, Gordon, & Gu, 2007) to guide and assess effective treatment for violent offending. As well, the VRS has been used to measure treatment change in a study of cognition in male offenders in New Zealand (Polaschek, Bell, Calvert, & Takarangi, 2010) and to described patients on a Dangerous and Severe Personality Disorder unit, without predicting outcome (Sheldon & Krishnan, 2009). The current

study will be the first to examine the VRS's ability to predict recidivism in a sample of female offenders.

1.7 Guiding Principles for Treatment of Female Offenders

Beyond examining the effectiveness of risk assessment instruments and additional gender-informed variables to predict recidivism, the assessment of criminogenic need areas also provides guidance for targeting interventions, so to reduce institutional misconduct and to prevent reoffence, as well as to aid in emotional and personal self-improvement (Bonta et al., 1995). In support of this position, CSC has moved towards a more women-centred approach with *Creating Choices* program (Task force on Federally Sentenced Women, 1990). *Creating Choices* attempts to better address the needs of female offenders by empowering women, helping them to make meaningful and responsible choices, treating them with respect and dignity, providing them with a supportive environment, and taking shared responsibility for our society and the impact is has on these women so to allow them take responsibility for their past, present and future situations.

The treatments available for female offenders are diverse and are targeted on needs beyond those currently identified as criminogenic needs, with the understanding that responsivity issues must also be addressed. For example, the majority of federally sentenced women are survivors of abuse and trauma and two thirds of these women have shown interest in voluntary programming or counseling to work through these issues (Federally Sentenced Women Program, 1994; Fortin, 2004). Other interventions include living skills programs, literacy and continuous learning programs, and substance abuse programs, sex offender therapy for women, reasoning and rehabilitation program (revised cognitive skills training), anger and emotion management program, mental health programs (e.g., Dialectical Behaviour Therapy, psychosocial rehabilitation), education programs, employment and employability programs, parenting skills programs, mother-child program, community integration program, choosing health in prison program, peer support program, leisure education program, lifeline program, spiritual and spirituality services, canine program, recreational therapy, horticulture program, arts and crafts, and aboriginal programming (Federally Sentenced Women Program, 1994; Fortin, 2004).

As reviewed above, Dowden and Andrews (1999) examined the effectiveness of correctional treatment for female offenders and found when it adhered to the same three psychological principles of risk, need and responsivity, used to evaluate male treatment

programs; it was most likely to result in reductions in recidivism. If the gender-informed need areas (i.e., history of victimization, economic pressures, suicide attempts and self-harm behaviour, and childcare responsibility) are not found to be directly related to recidivism (i.e., criminogenic needs) in the current study, they may still be interfering with the individual's ability to benefit from treatment, and thus impeding or minimizing the impact on recidivism (i.e., responsivity issues).

1.8 The Purpose of the Current Program of Research

This study compared four common actuarial risk assessment instruments to determine which best predicted general and violent recidivism and institutional misconduct (antisocial behaviour in a more immediate form). In addition, gender-informed variables (i.e., history of victimization, childcare responsibility, economic pressures and self-injury and suicidality) were explored as potential criminogenic needs by examining their relationships with general and violent recidivism and institutional misconduct. Next, composites of these gender informed variables suggested from the pathways and gender-responsive literature (e.g., being a single parent, the combination of abuse history and substance abuse) were also related to a variety of recidivism outcomes to determine if they are more useful in predicting outcome when examined together. Finally, the incremental validity of these gender informed variables and composites was explored over and above the risk assessment instruments.

The current study could lead to the validation of several popular risk assessment instruments for use with federal female offenders and contribute to the general body of empirical research on women offenders. While replication would be required, this comparison of common risk assessment instruments could form a strong foundation for actuarial risk assessment becoming part of the common practice of evaluating federal female offenders. Improving risk assessment in this population will allow clinicians and other correctional practitioners to make decisions about these offenders more accurately, and the exploration of the gender-informed variables could help clarify the risk/needs of federal female offenders that may have gone undetected. Alternately, if gender informed needs expected to be directly predictive of recidivism are not, but are prevalent for female offenders, they may suggest areas to explore in the future as specific responsivity factors. As well, further clarifying the applicability of the risk, need and responsivity principles of the PIC-R theory of criminal behaviour will help elucidate the utility of this theory for federal female offenders specifically, and female offenders in general.

1.8.1 Hypotheses.

As noted above, the purpose of this study was to examine the efficacy of traditional needs/risk assessment instruments for female federal offenders and whether the predictive value of these instruments could be increased by adding gender-informed variables. Specific hypotheses are provided below.

1.8.1.1 Psychometric properties.

Each of the traditional needs/risk assessment instruments was expected to have good internal consistency and inter-rater reliability.

1.8.1.2 Predictive validity.

The traditional needs/risk assessment instruments were expected to have significant predictive validity for general and violent recidivism, institutional misconduct, and severity of reoffence. Significant correlations between risk instruments and outcome, and Receiver/Relative Operating Characteristics (ROC) analyses with Areas Under the Curve (AUC values) significantly greater than .50 were expected (Rice & Harris, 1995). Specifically, the LSI-R and LS/CMI were expected to have the strongest relationships with general recidivism, revocation, and institutional misconduct, as evidenced by significantly stronger correlation coefficients as compared to the other traditional need/risk assessment instruments. It was hypothesized that the LSI-R and LS/CMI would be the best predictors of institutional misconduct and general recidivism, as they have been validated on provincial female offenders (Rettinger, 1998; Rettinger & Andrews, 2010) and have outperformed other actuarial risk assessment instruments with male offenders (Gendreau et al., 1996).

Further, the VRS was expected to have the strongest relationships with violent recidivism and offence severity, as evidenced by significantly stronger correlation coefficients as compared to the other traditional need/risk assessment instruments. The VRS has yet to be compared to other actuarial instruments making it difficult to predict its relative performance. However, given the fact that it includes both dynamic and static factors like the LSI-R/LS/CMI, yet is specifically aimed at violence, it was expected to be the best predictor of violent recidivism.

1.8.1.3 Predictive validity of risk levels.

Women in the high risk groups on each of the traditional needs/risk assessment instruments were predicted to have higher and faster rates of general and violent recidivism than those with lower scores, as illustrated by survival analysis. Further, significant positive correlations are also expected between the risk levels of each traditional needs/risk assessment instrument and outcome, except for the SIR-R1 scale where significant negative relationships are expected.

1.8.1.4. Incremental predictive validity.

The gender informed variables were expected to have significant predictive accuracy for violent and general recidivism, revocation, institutional misconduct, and severity of reoffence, as exemplified by significant correlations and AUCs between each relevant variable or composite and outcome. Because some of these variables seem to overlap with related items on some of the risk assessment instruments (e.g. number of children/dependents on SIR scale, marital status on SIR, LSI-R, LS/CMI, etc.), duplicates were not expected to add to these instruments in the prediction of general or violent recidivism.

However, their composites (e.g., marital status and number of dependents can be combined to look at the effect of single parenting) were expected to be especially useful predictors of recidivism on their own when compared directly with outcome, as well as, when added to the risk assessment instruments in which they are not contained. Further, these composites were expected to increase the predictive accuracy of the traditional need/risk assessment instruments for general and violent recidivism, revocation, institutional misconduct, and severity of reoffence.

THE PRESENT STUDY

2.1 Methods

2.1.1 Participants.

2.1.1.1 Inclusion criteria.

One hundred and one female federal offenders were included in this study. These women were randomly selected from a national cohort of all federal female offenders with warrant expiry dates (WED) between January and December 2002 (N=282). This allowed for an approximately seven-year timeframe for a post-release follow-up period. This cohort was further reduced (N=268) by ensuring these women were admitted to a Canadian federal correctional institution/CSC on or following January 1, 1995 to ensure that their files were available in electronic form through the Offender Management System (OMS). Note, this cohort, given the parameters described above, did not include any women with indeterminate sentences. Any female offenders with sentences less than two years (likely provincial offenders supervised by CSC due to an exchange of service provincial agreement) were also excluded (eliminating another 53 cases). Moreover, two additional cases were eliminated as the women were deceased prior to their WED, and one case as the woman was unlawfully at large (UAL).

With the remaining cohort of 212 women, files were screened to ensure minimal file completeness likely to allow for adequate to strong ratings (i.e. information about the individual's risk/need domains as typically assessed by CSC, and the circumstances of their index offence). More specifically, each file must have contained at least an Intake Assessment report and a Criminal Profile report or a Correctional Plan which also included initial intake assessment information, or comparable reports (e.g., a thorough Preliminary Assessment report). This screening procedure yielded 201 viable cases for review.

Further, 134 cases were randomly selected from the 201 viable cases for review. This was done to reach the targeted sample size of 100 participants with adequate files for review. An additional 33 cases were excluded due to having insufficient reports in the assessment period to allow an adequate to strong rating (e.g., some files had no documents in the first year of their sentence), being written exclusively in French, the offender being deported and thus follow up was not possible, or the OMS record no longer existing at the time of file review. The final sample included 101 rated files.

2.1.1.2 Representativeness of the sample.

To ensure that the sample well represented the cohort from which it was drawn, basic demographic information was compared for all of the women who had received sentences two year or longer (N=229). Of that cohort, 10.9% were North American (Aboriginal) and an additional 5.2% identified as Metis and 0.9% as Inuit, while the majority was Caucasian (57.6%). African Canadians comprised the next largest ethnic group at 14.4% of the cohort. Of the 101 female offenders sampled in this study, 15.8% were North American (Aboriginal) and an additional 6.9% identified as Metis, while the majority were Caucasian (57.4%). African Canadians again comprised the next largest ethnic group at 11.9% of the sample. Chi Square analyses on ethnicity of those in the cohort who were randomly selected to be in the sample and were not selected yielded $\chi^2(12, N = 229) = 20.03$, $p = 0.067$ suggesting that the sample was not significantly different from the remaining cohort with regard to ethnicity.

Further, the mean age of the cohort at the time of assessment was 34.24 years (SD=9.22), while the mean age of the sample at the time of assessment was 34.39 years (SD=8.34). A t-test analysis suggests that the sample did not significantly differ from the cohort with regard to age, $t(227) = 0.209$, $p = 0.834$.

The sample did differ from the cohort with regard to sentence length of the index offence, $t(162.25) = -3.415$, $p = 0.001$, not unexpectedly, as the sampling procedures eliminated any women with sentences beginning prior to 1995 (to ensure that full electronic files were available for review). There were 13 cases with sentences commencing prior to 1995 in the original cohort with sentences ranging from just over eight years up to almost 30 years. The mean sentence length (in days) at sentence commencement for the sample was 1057.01 (SD=453.94) and for the cohort it was 1494.62 (SD= 1340.40).

2.1.1.3 Additional sample characteristics.

Fifty three percent of the women in the sample were married or in common law relationships at the time of index offence. Sixty eight percent were mothers. While all of the women in the sample had committed an offence leading to a sentence of at least two years (which was the most frequent index sentence), the average sentence length was 1081.47 days (SD=539.319; note this sentence length is slightly greater than that reported above when comparing sample and cohort as this sentence length is recorded as of the date of assessment and

some women had outstanding charges that were dealt with during the assessment period affecting their index sentence length).

Follow up time (from the latter of first release and the end of the assessment period) ranged from 780.75 to 4487 days, with a mean follow up time of 3021.53 (SD=591.78) days. These numbers include data from six participants who died during the follow up period (but, who were followed for at least two years post release), and were retained in the study due to the small sample. First release type was day parole for 84% of the sample, full parole for 4% and statutory release for the remainder of the sample (13%). Forty eight percent of the sample was revoked at least once during their index sentence. Fifty seven percent were charged with institutional misconduct (and 32% with serious misconduct) at some point in their index sentence, while 42% continued to engage in misconduct after the assessment period. Finally, 52% of the sample recidivated (resulting in a new reconviction for any offence), while 11% recidivated violently (resulting in a new conviction for an offence against a person causing physical or psychological harm).

2.1.2 Materials.

2.1.2.1 Offender Management System (OMS).

File reviews of the female offenders were completed using OMS. OMS is a secure system of CSC's electronic offender files. The OMS files include documents such as the Offender Intake Assessment and the Criminal Profile Report or Correctional Plan that aim to gather a variety of interpersonal and personal influences present in an offender's life, exploring both need and risk. Needs ratings generally focus on the extent to which difficulties in the offender's life provide insight into their life histories and lifestyles before being incarcerated, as well as their programming requirements (Dell & Boe, 2000). Risk ratings assess the probability of reoffence if the identified needs are not thoroughly addressed (Dell & Boe, 2000).

2.1.2.2 Level of Service Inventory-Revised (LSI-R).

The LSI-R (Andrews & Bonta, 1995) is an instrument designed to explore an individual's risk for criminal behaviour and their treatment needs. The instrument includes 54 items grouped into 10 risk/needs subscales: Criminal History, Education/Employment, Financial, Family/Marital, Accommodation, Leisure/Recreation, Companions, Alcohol/Drug problem, Emotional/Personal, and Attitudes/Orientation. Items are scored as present or absent, or on a scale from 0 to 3, where 0 represents "a very unsatisfactory situation with a very clear and strong

need for improvement" and 3 represents "a satisfactory situation with no need for improvement" (Andrews & Bonta, 1995). Items scored on the 0 to 3 scale are then converted to present or absent scores, where 0 and 1 become present and 2 and 3 become absent. The items are then summed to yield a total score. The maximum possible score is 54, and represents the highest risk for reoffence and highest need for treatment.

The LSI-R has been widely validated on a variety of samples of male and female offenders from around the world, including Canadian federal and provincial offenders. The manual reports interrater reliability, as defined by an agreement statistic (r), ranging from .84 to .99 (Andrews & Bonta, 1995) across a variety of samples. Specifically with regard to female offenders, Rettinger (1998) found a high interrater reliability estimate at .92 for provincial offenders. Measures of internal consistency for the LSI-R typically yield alpha values in the .70s and has been shown to have good predictive validity with probationers and inmates on a variety of outcomes including reconviction, institutional misconduct, and parole outcomes (Andrews & Bonta, 1995). Specifically with regard to female offenders, Coulson and colleagues (1996) and Rettinger (1998) found the LSI-R predicted reconviction in samples of provincial offenders very well (r=.51 and r=.58, respectively).

2.1.2.3 Level of Service / Case Management Inventory (LS/CMI).

The LS/CMI (Andrews, et al., 2004; Formally, the Level of Service Inventory-Ontario Revision, LSI-OR) retained the same scoring system as the LSI-R (described above), however it contains only 43 items and 8 risk/needs subscales, having eliminated Accommodation, Personal/Emotional and Financial and adding Antisocial Pattern. These 43 items are found in section A of the instrument. This instrument also allows promotive factors (strengths), beyond the absence of risk, to be considered, as well as, the specific need, risk and responsivity factors of the individual (compiled in section B, e.g. poor social skills) to be incorporated into their total assessment using a clinical override. For the purpose of this study overrides will not be applied and therefore only the risk rating derived from section A will be evaluated. As with the LSI-R, a higher score represents a greater risk for recidivism and greater treatment needs.

As an extension of the LSI-R, the LS/CMI has also been well validated and its reliability thoroughly examined. Andrews, Bonta, and Wormith (2004) report internal consistency (Cronbach's alpha) for female offenders from their North American normative data to be between .91 and .94. Studies examining interrater reliability on the LS/CMI have been

confounded with test retest reliability and only results on a small adult sample are provided in the manual with r=.88 (Andrews, et al., 2004). With regard to validity, again the LS/CMI has been extensively examined in a variety of samples. Looking specifically at female offenders, Rettinger (1998) found the LSI-OR predicted reconviction in samples of provincial offenders very well (r=.59). Further examination of Rettinger's results reveals that the LS/CMI total score has a correlation with general recidivism of .63 for her total sample and .44 for violent recidivism (Andrews, et al., 2004) again suggesting strong predictive utility of this instrument for provincial female offenders.

2.1.2.4 General Statistical Information on Recidivism Revised (SIR-R scale).

The SIR-R scale (Nafekh & Motiuk, 2002, adapted from the General Statistical Information on Recidivism, GSIR, Nuffield, 1982) is an actuarial instrument used by CSC and the National Parole Board of Canada with non-aboriginal male offenders to assess the risk of recidivism providing probability estimates for reoffending within three years of release (five groupings ranging from 80% of offenders predicted to succeed to 33% of offenders predicted to succeed). The scale includes 15 items examining criminal conduct (e.g., age of first arrest and type of offence) as well as demographic variables (e.g., marital status). Items are reverse scored, so that the lower the summed score of all items the greater the probability of recidivism, and ranges from -30 (high risk) to +27 (low risk). The revised scale (SIR-R1) differs from the original with regard to this reverse scoring, as also reflects other changes to improve face validity (scoring no longer reflects lowered risk for individuals with a previous sexual offence compared to those without) and changes in legislation (including the *Young Offender's Act* and changed definitions for mandatory supervision and statutory release; Nafekh & Motiuk, 2002).

Numerous studies have shown good predictive validity for the GSIR/SIR-R1 scale (Nuffield, 1982; Bonta, Harmann, Hann, & Cormier 1996; see Nafekh & Motiuk, 2002 for a review). Wormith and Goldstone (1984) also found some support for male offenders with better scores on an earlier version of the GSIR achieving early release and being successful during community supervision. While past research on the GSIR/SIR-R1 scale has reflected difficulty predicting violent recidivism, this may be attributed to the low base rate of this outcome. Nuffield (1982) did not find any of the GSIR items correlated with violent recidivism. However, Bonta, Harmann, Hann, and Cormier (1996) showed some improvement in prediction over chance with a sample of men with a similar violent recidivism rate using ROCs and thereby

avoiding base rate influences (AUC = .64 or .65, depending on their definition of violence). More recently, Kroner and Loza (2001) found the GSIR to be a strong predictor of violent recidivism (AUG = .74).

Studies examining the utility of the SIR scale for female offenders have provided mixed support for the GSIR/SIR-R1 scale's use with this population. Bonta and colleagues (1995) did not find a linear progression from one risk grouping to the next, although did find the GSIR (based on similar file ratings) was mildly predictive of general recidivism. Nafekh and Motiuk (2002) explored a SIR proxy scale on 342 federal female offenders using information from the Offender Intake Assessment indicators to approximate the SIR-R1 items. The AUC was .767 for general recidivism and .725 for violent recidivism and they suggested that their SIR proxy was highly correlated and just as effective of the SIR-R1 scale. Nafekh and Motiuk (2002) also indicate that the SIR proxy accurately distinguishes between the five risk groups, although they suggest that these groups should be redeveloped on female samples in a method similar to the original scale construction with men.

2.1.2.5 The Violence Risk Scale (VRS).

The VRS (Wong & Gordon, 1998 – 2003) rates six static and 20 dynamic factors pre-treatment to assess an individual's level of risk and needs. The static factors include, for example, age at first offence, current age and stability of upbringing. The dynamic factors include, for example, violent lifestyle, criminal personality, and criminal attitudes. Each factor is rated from 0 to 3 where a lower score represents less risk or fewer needs. The sum of the total static factors and the total dynamic factors represents an individual's level of risk pre-treatment. The maximum pre-treatment score would be 78 and represent the greatest risk and the most treatment needs. When the efficacy of treatment is to be examined, each of the dynamic factors should also have one of four stages of change ratings (ranging from contemplation or precontemplation to maintenance). A post-treatment assessment can then be completed by indicating the stage of change the offender achieved post-treatment and by examining their progression through the stages for each of the dynamic factors relative to their pre-treatment ratings. For the purpose of this study only, pre-treatment ratings will be conducted.

The initial version of the VRS, the VRS-Experimental Version1, was validated on a sample of provincial and male offenders (Gordon, 1998). There is currently little data on the use of the VRS with female offenders. Wong and Gordon (2006), in a review of the validity and

reliability of the VRS to date, report high interrater reliability (with ICC=.92 and .97, and r=.87) with male offenders. With regard to predicitive validity, correlations are good between VRS total score and general recidivism ranging from .37 to .43, and with violent recidivism ranging from .28 to .40 (static and dynamic subtotals yielded similar correlations with general recidivism, ranging from .33 to .42 and .35 to .40, respectively, and with violent recidivism ranging from .21 to .31 and .28 to .40, respectively). Dolan, Fullam, Logan, and Davies (2008) included 11 women in their total sample of 147 forensic patients and reported that conducting their analyses with and without the female participants did not markedly affect the predictive validity scores of the VRS and institutional violence.

Stockdale (2008) examined a related instrument, the Violence Risk Scale-Youth Version (VRS-YV; Lewis, Wong, & Gordon, 2004), and included female young offenders in her study. High interrater reliability was observed for the VRS-YV static (*ICC* = .87), dynamic (*ICC* = .89), and total (*ICC* = .90) scores on her overall samples including males and females. Explorations of predictive validity for the female youths specifically yielded correlations between the total scores and general recidivism of .36 and .41 with violent recidivism (static subtotals correlated .18 with general recidivism and .32 with violent recidivism, while dynamic subscale scores correlated with general recidivism .39 and .40 with violent recidivism). This suggests that the parent instrument, the VRS, may be promising for adult female offenders as well.

2.1.2.6 Gender informed variables.

These variables represent some of the constructs thought to be relevant to the needs/risk appraisal of female offenders. Some of these variables may be similar to items on the other risk assessment instruments; however they are believed to be of significant relevance to risk assessment and were collected separately. These variables include *childcare responsibility* variables (the number of children/dependents residing with her when not incarcerated and in total; whether she has a current spouse/common law partner), *economic stress* (i.e., a history of illegal financial support, a history of social assistance use), *history of victimization* variables [the self-reported presence, frequency (one specific incident, several specific incidents, enduring abuse over weeks, months, or years) and type (physical, emotional, sexual) of childhood and adult abuse], and the *presence of self harm and/or suicide attempts*. Note, for the purpose of this study "emotional abuse" was rated as being present if the woman's file noted examples of such

abuse specifically (e.g., being degraded, name calling, belittling, etc.). It is acknowledged that there are many ways to operationalize emotional abuse. Individuals who were physically or sexually abused may also have experienced emotional abuse in some way during those events (whether this was noted specifically or not), but it can also occur independently (Claussen & Crittenden, 1991) and the goal in this study was to examine any effect of emotional abuse as acknowledged and noted specifically by these women. Further, emotional abuse has been more strongly correlated with experiences of neglect than severity of physical abuse (Claussen & Crittenden, 1991). Finally, self harm and suicidal behavior were not distinguished from one another in this study as the poor quality of the file information on this topic (e.g., limited or no detail of such events, differing definitions of self harm and suicide by file authors) made it difficult or impossible to obtain adequate information to make such a differentiation.

2.1.2.7 Recidivism.

General recidivism was defined as an incident involving a female offender being convicted for an additional offence post-release and post-assessment period. These offences included nonviolent crimes such as obstruction, theft, fail to comply, attempted robbery, fraud, uttering threats, break and enter, communication for the purpose of prostitution, driving under the influence, and possession of scheduled substance, as well as violent crimes. *Violent recidivism* is defined by an incident involving the female offender being convicted for an additional offence post-release and post-assessment period for actions against another person resulting in physical or psychological harm. These offences included violent crimes like assault, assault with a weapon, assault causing bodily harm, robbery and aggravated assault. Although not present in the current study homicide related offences, kidnapping, abduction, unlawful confinement, dangerous weapon use, criminal negligence causing harm, arson and any driving offence causing bodily harm or death would have been classified as violent offences.

In addition to the presence and absence of each type of recidivism, the frequency of each type of recidivism was also collected. Time to first reoffence for general and violent recidivism was also calculated. This was defined as number of days post release or post-assessment period (which ever was later) until first reoffence (general or violent, respectively). Additionally, number of revocations and time to first revocation were also tracked.

Further, the *severity of reoffence* was included as an outcome measure. This was defined in a number of ways. First, severity was defined as the longest incarcerated sentence length

earned for a new offence(s) in days. Secondly, to allow for accurate representations of sentence length when multiple convictions were received (on the same or multiple sentencing dates) during the follow up period, the overall sum of aggregate sentences for each incarceration, probation, and conditional sentences were calculated separately. To do so, concurrent and consecutive sentences were combined appropriately to reflect actual time sentenced, and this was achieved by following rules typically employed by sentence managers/administrators (those tasked with computing official sentence lengths for offenders and correctional agencies). Several manuals were referenced to aid in these computations including: *Sentence Calculation: A Handbook for Judges, Lawyers and Correctional Officials: Third Edition,* (Public Safety and Emergency Preparedness Canada, 2005); *Introduction to Provincial Jurisdiction Adult Sentence Calculation: A Handbook for Criminal Justice Professionals*, (Saskatchewan Corrections & Public Safety, 2004); and *Conditional Sentence: Introductory Sentence Calculation Handbook for New Sentence Administrators*, (Saskatchewan Corrections & Public Safety, n.d.). Sentence length has been shown in past research to be a valid representation of offence severity (Belanger, 2001).

Finally, two validated measures of offence severity were also coded from recidivism data. These measures included Offence Categories Ranked by Seriousness (OCRS) which was developed by the Research Department of the Ministry of Correctional Services of Ontario in 1982 (Stasiuk, Winter & Nixon, 1996). This scale grouped offences similar in sentence type and duration. The groups were rank ordered based on the average sentence length imposed by Ontario courts for these kinds of offences as based on the sentences given to 60,000 offenders in one year. Offences that did not exist at the time that the scale was developed were subsequently added based on their similarity to other included offences (e.g., Infanticide and Murder Two were grouped with the other homicide and related offences as the most severe). The rankings range from 1 to 24, with the lowest score being most severe (i.e., homicide and related offences). Cases where offenders did not recidivate were coded as 25.

The second measure of severity, the Severity Index, was developed by the Canadian Centre for Justice Statistics and used when presenting the results of their Uniform Crime Reporting Survey (Statistics Canada, March 2006) as well as their Female Offenders in Canada issue of Juristat (Statistics Canada, 2008). This measure rank orders offences based on the maximum penalty prescribed by the Criminal Code of Canada. Further, all violations against the

person are rated as more severe than non-violent offences and federal statutes are ranked as more severe than provincial statues. The rankings range from 0 (i.e., no offences committed) to 13 (i.e., violations against the person with maximum penalties of 25 years), with the highest score being most severe.

2.1.2.8 Institutional misconduct.

Given that the goal of needs/risk appraisal is not focused solely on predicting criminal behaviour upon release, institutional misconduct was also be evaluated. Institutional misconduct is a more immediate form of reoffence shown to have similar predictors as recidivism upon release (Loucks & Zamble, 2000) and to be predictive of recidivism upon release (Bonta et al., 1995), as well. Institutional misconduct is defined by the presence or absence of the incidents of antisocial behaviours while incarcerated. The number and types of these antisocial behaviours was recorded where possible, as represented by the frequency of institutional misconduct (major and minor) for the complete index sentence duration, and for the index sentence following the assessment period only.

2.1.3 Procedure.

2.1.3.1 File reviews and ratings.

While most risk assessment instruments recommend interviewing offenders directly, given the difficulties in accessing female offenders across the country, the current study depended upon file review alone. Grann and colleagues (2000) completed the VRAG using file review only and were still able to significantly predict reconvictions within two years of release in mentally disordered male offenders. They further explained that is it feasible to assume that file review alone is an underestimation of actual predictive value of the instrument as compared to having trained clinicians performing live evaluations of the offenders. Similar findings exist for the PCL-R, suggesting that estimates of psychopathy in men based on file review alone are also more conservative (Wong, 1988).

The study involved a comprehensive file review of female federal offender OMS files from across Canada involving the collection of data to complete the LSI-R, LS/CMI, SIR-R scale, and VRS, and information on childcare responsibilities, economic stress, history of victimization and self-harm and suicidality. To ensure that the current study was consistent with past research and common CSC risk assessment practice, and to ensure that assessment data were not confounded by treatment effects; intake risk assessments, not pre-release risk

assessment, were collected for each female offender. Risk ratings were based primarily on the first one year of each female offender's sentence. The title and date of each report is listed in OMS in table form, and only reports within the specified time frame were accessed. In addition, each sentence receives its own screen listing the available reports in OMS, therefore reports for later sentences were not displayed along with those being rated, and thus the rater remained blind to the offenders' outcomes. Completing risk assessment early in the sentence meant that the ratings were also conservative predictors of recidivism upon release, as they did not reflect the changes in dynamic risk due to treatment. In addition, to ensure that the raters were blind to outcome, the recidivism data was collected separately.

Of note, there were some methodological decisions made about rating female offenders that differed from the typical procedures used for file ratings for research purposes. Namely, all documents available in the first year were reviewed, not just key documents, as after reviewing a subset of files, it was apparent that pertinent information was often noted in some documents that did not appear to be reflected in amalgamated reports (e.g., Correctional Plans). Further, given that there was no opportunity to interview the offenders, reading as many authors' perspectives on the offenders allowed for a more comprehensive reflection of their varied opinions of the women's attitudes and behaviours. Also, reviewing entire files likely mirrors best practices for risk assessments most accurately.

Participant's scores were included with missing data, as permitted, and scores were prorated, where applicable, as prescribed in each risk assessment manual. The VRS manual only outlined how to prorate total scores and was silent on the two subscale scores. Therefore, these scores were prorated in a similar fashion to the total score (subscale score X number of subscale items / number of subscale items – number of omitted items) when they were explored independently (as in previous research has done). Thus, summing prorated subscale scores does not yield the prorated total score.

2.1.3.2 Interrater reliability.

Interrater reliability training included formalized training on the LS/CMI and VRS as provided by the authors and official trainer for these instruments. Given the similarity of the LS/CMI and the LSI-R, the primary researcher provided supplemental training on the LSI-R to the interrater, as well as training on the SIR scale (with support from other CSC researchers). Sample cases from the instruments' training materials were coded prior to training using files

from the female offender sample. The primary researcher and interrater compared ratings at various points when completing interrater reliability to ensure agreement and to clarify procedures described in the coding manuals for commonalities in the sample (e.g., how to rate VRS items for women with no history of violent offending). Measures of agreement were also calculated and compared to help in determining when training was sufficient.

Generally, inconsistencies in files between authors were resolved using the raters' judgment about which information or perspective was most frequently and convincingly reported (e.g., accounts with more details provided or reports where it was clear the author was the original source for this information not earlier file information were less likely to be ambiguous or misunderstood). For some items (e.g., attitudes), it was decided that additional training was unlikely to improve interrater reliability as file information was often inconsistent in this regard. These items relied heavily on the authors' opinions of the women, which could vary significantly between authors, and it was difficult to prescribe how raters should objectively weight these differences of opinion. These constructs were felt to be less reliable and not reflective of inadequate training or operational definitions. Note, most other discrepancies in file information were reconciled by consensus by the raters. Additional guidance was provided about how to address differing accounts of abuse. It was decided if at any point a woman acknowledged experiencing abuse it was rated as occurring, even if other authors denied any such events. This is because self-report may have greatly depended upon the woman's comfort with an author, and it was felt that giving the women the benefit of the doubt about such experiences was reasonable despite the risk of dishonesty.

2.1.4 Statistical analyses.

2.1.4.1 Psychometric properties.

The interrater reliability of the traditional needs/risk assessment instruments and gender informed variables were examined on a random selection of 15 cases collected and coded independently by a second rater. Interrater reliability was evaluated by calculating the intraclass correlation coefficient (ICC). Two-way, mixed, single rating ICCs were explored, with absolute values reported for individual items, while consistency values were reported for subtotal and total scores, consistent with best practices (McGraw & Wong, 1996; Nichols, 1998; Shrout & Fleiss, 1979). The ICC not only examines the correlation among the ratings, but also considers potentially different anchor points among raters as it is a measure of agreement, not association

(Bartko, 1991; McGraw & Wong, 1996). In general, reliability coefficients around .70 or higher are considered to be acceptable, while values of .80 are typically regarded as moderate to high levels of reliability (Murphy & Davidshofer, 2005).

Internal consistency was examined using Cronbach's alpha. Note, as Green and Lissitz, and Muliak (1977) clarify internal consistency implies interrelatedness of items, but not necessarily a unidimensional construct. Cortina (1993) suggests caution when using alpha alone to determine if a scale reflect a unidimensional construct, especially when a scale has a larger number of items, as higher alpha values would be expected even with lower item intercorrelations and factor analytic techniques should be used to clarify the dimensions of a scale. This may be especially relevant to the current examination, as by design these scales (and the gender informed variables) measure multiple items in a number of domains or subconstructs thought to be related to the construct of risk for reoffence which is acknowledged to be heterogeneous.

2.1.4.2 Predictive validity.

First, correlations were conducted on each of the total scores of the actuarial risk instruments with general recidivism, violent recidivism, revocations, and incidents of institutional misconduct, to examine the strength and direction of these relationships. Correlations of the actuarial risk instruments and gender-informed variables and severity of recidivism were also examined.

Cohen (1992) also suggests that correlations in the .10 range represent small effect sizes, in the .30 range medium effects sizes and in the .50 range large effect sizes. However, Meyer and colleagues (2001) remind researchers should be "satisfied" with correlations at the .10 to .19 range, "pleased" with correlations in the .20 to .39 range, and "rejoice" at values that are higher given the difficulty of consistently achieving uncorrected univariate correlations greater than .30 in real world settings.

Bonferroni adjustments or other methods of adjusting statistical significance levels (i.e., p-values) to address the family wise error rate were not employed in the current study. Rothman (1990) cautioned against making adjustments for multiple comparisons as he argued that data being examined are not random numbers, but represent actual observations governed by natural laws and suggests that researchers should be less reluctant to explore leads that may be incorrect to avoid missing potentially important findings. He highlighted that there is no formula

that can replace the critical evaluation of results observed. Further, Perneger (1998, p. 1236) suggested that "Bonferroni adjustments are, at best, unnecessary and, at worst, deleterious to sound statistical inference." This is because such adjustments create new methodological problems (e.g., increases the risk of a type II error; which may be especially problematic in new areas of study), and the Bonferroni method implies that all null hypotheses are true simultaneously (Pernerger, 1998) which is not the focus of the current investigation. That is, the goal of the current study is not to say whether all variables are related to all outcomes or not, which a Bonferroni correction would allow, but instead to examine each variable in its own right. Instead, Pernerger (1998) suggested describing statistical methods and rationales as well as possible interpretations to help the reader make reasonable conclusions without Bonferroni adjustments. There are a growing number of researchers who have supported this position (Cabin & Mitchell, 2000; Moran, 2003; O'Keefe, 2003). In the current study, the examination of multiple definitions of related variables also provides some form of within study replication for the findings.

Further, confidence intervals were also provided where possible as they are a good alternative to statistical significance corrections (Garamszegi, 2006). Thus, these correlational relationships were also illustrated with ROC analyses. These analyses produce Areas under the Curve (AUCs) which, for the purpose of this study, represent the probability that a randomly selected recidivist will have a higher risk score than a randomly selected nonrecidivist (Hanley & McNeil, 1982). Moreover, AUCs are not affected by base rates or selection ratios, which is especially important when predicting violent recidivism (Mossman, 1994; Quinsey, Harris, Rice, & Cormier, 2006) given that only 11% of this sample recidivated violently. AUC values at .50 represent chance prediction, while those at 1 represent perfect prediction (Mossman, 1994; Quinsey, Harris, Rice, & Cormier, 2006). Rice and Harris (1995) proposed that AUCs of 0.64 to 0.70 represent a moderate effect size and AUCs of 0.71 and higher represent a large effect size.

2.1.4.3 Predictive validity of risk levels.

Survival analyses allow the examination of time to an event, as well as the rate of this event relative to a comparison group (Luke & Homan, 1998). In this study, survival analyses allowed the cumulative percentage of offenders who did not recidivate (that is, those who have "survived") to be tracked during the available follow up period. Further, by grouping offenders based on their risk ratings their relative performance could be compared. While there are a

number of nonparametric statistics that can be used to ensure that the groups (in this case actuarial risk levels) actually differ in their survival probabilities, with a small sample size or longer follow up time Luke and Homan (1998) suggest interpreting the generalized Wilcoxon or Tarone-Ware tests. For, these tests earlier observations are weighted more because there are more cases remaining in the risk pool and the tails of these curves can less reliable.

To perform survival analyses, the time women were released in the community was compared. For recidivists this was defined as time to first reoffence (for general or violent reoffence, respectively) from the latter of first release date or the end of the assessment period. This was compared to follow up time for nonrecidivists, which also began with the later of release or post-assessment period and ended with the date that all recidivism data were collected. Offenders were compared grouping them based on their risk scores as outlined in each instrument's manual or by truncating the data into near equal thirds, as was done with the VRS, to create low, medium and high risk groups. Specifically, for the SIR-R, scores above and including 6 were classified as low risk, 5 to 1 as low to moderate risk, 0 to -4 as moderate risk, -5 to -9 as moderate to high risk, and -10 and lower as high risk. For the LSI-R, scores from 0 to 13 were low risk, 14 to 23 were low/moderate risk, 24 to 33 were moderate risk, 34 to 40 were medium/high risk and scores of 41 and above were high risk. For the LS/CMI, there were no scores in the very low risk level, scores from 6 to 10 were low risk, 11 to 19 were medium risk, 20 to 29 were high risk, and over 30 were very high risk. Finally, for the VRS low risk was scores up to and including 12, medium risk was 13 to 36, and high risk was 37 and up.

2.1.4.4. Incremental predictive validity.

To explore the relationship between recidivism (general and violent recidivism, as well as revocation and institutional misconduct) and the gender informed variables (childcare responsibility, presence of spouse or common-law, presence substance abuse, substance abuse intensity, illegal financial support, history of social assistance, self harm/suicide, child physical abuse, child sexual abuse, child emotional abuse, adult physical abuse, adult sexual abuse, and adult emotional abuse) were correlated. As was done for each of the risk assessment instrument, ROC analyses were also conducted for each gender informed variable with each outcome variable.

Next, a subset of the gender-informed variables was combined to look at the gender-informed composites that may help clarify some of the inconsistencies found in past research.

Namely, variables looking at presence of spouse or common law and number of dependents were combined to explore whether being a single parent is predictive of recidivism (general, violent, institutional misconduct) for female offenders in addition to whether marital status or childcare responsibilities alone were predictive. This was explored using a series of correlations.

Secondly, after examining the past victimization variables and substance abuse alone, each relevant type of victimization was combined with substance abuse to see whether the combination of victimization and substance abuse was predictive of recidivism as suggested by McClellan, and colleagues (1997). This was again explored using a series of correlations.

Next a series of backward elimination stepwise logistic regression analyses was used to determine whether the gender-informed variables and their composites predicted outcome over and above the risk assessment instruments alone. Specifically, a risk assessment instrument was entered into its own block, followed by the gender-informed variables or composites in a second block in stepwise form to determine if they increased the effect size of the model over and above that of the instrument alone. Moreover, for this study, the group of gender informed variables and the group of composites were explored initially in separate logistic regression analyses to avoid increasing the error variance by including a number of highly related variables in the same analysis (recall, the composites were derived from the gender informed variables). Once individual gender informed variables and composites were identified in separate stepwise logistic regression analyses as contributing significantly to the prediction of outcome over and above the risk assessment instrument being examined, only these gender informed variables and composites were combined in one analysis to derive a final solution representing the best combined prediction. Note, when the presence of substance abuse, abuse variables and their related composites were shown to significantly add to the predictive utility of their individual models, the presence of substance abuse with not included in the final combined model to avoid violating the independence of errors assumption for linear regression. This series of analyses was conducted separately for each risk assessment instrument

Although stepwise methods are criticized for being atheoretical, overly affected by random sampling variation and at risk of over or under-fitting the model, they are defensible for exploratory model building when previous research upon which to base a hypothesis is lacking (Field, 2009). Further, backward elimination, as compared to forward elimination, reduces the risk of eliminating a predictor involved in a suppressor effect, and thus reduces the risk of a type

two error (Field, 2009). The likelihood ratio statistic was used to examine the steps in the model to evaluate improvements in fit, as the Wald statistic can be unreliable and produced increase risk of a type 2 error also. The Wald statistic was used to examine the influence of an added variable or composite in predicting outcome within a given model. The *Exp (β)*, or exponentiated regression coefficient, is an odds ratio statistic reflecting the degree of change in risk of outcome based on a change in that predictor variable. When *Exp (β)* is greater than 1 this represents a positive relationship with outcome, while a value less than one represents a negative relationship. For example, an *Exp (β)* value of 2 suggests that for each one point increase in the value of the predictor variable (or the presence of that variable if it is dichotomous), the risk of the targeted outcome doubles. Similarly, with an *Exp (β)* value of .5 the risk of outcome is cut in half with a one point increase in the predictor variable. It is especially important that the confidence interval for the *Exp (β)* does not cross 1 or the interpretation of the variable (i.e., the direction of the relationship) becomes unreliable (Field, 2009). See Hosmer and Lemeshow (1989) for a detailed review of logistic regression.

3.1 Results

3.1.1 Data screening and basic descriptive statistics.

All missing data were within acceptable limits as prescribed by the risk assessment instruments. For the LSI-R, 28 cases had no missing items, 37 cases had one missing item, 21 cases had two missing items, 10 cases had three missing items, four cases had four missing items, and only one case had five missing items (the maximum allowable). For the LS/CMI, 42 cases had no missing items, 50 cases had one missing item, seven cases had two missing items, and two cases had three missing items (four missing items is the maximum allowable). The VRS describes missing data as either "not applicable" or "insufficient information." Often when a rating of "not applicable" was given it was often because the female offender in question did not have a history of violent offending or behavior. This rating was given 25 times in total (across all cases and all items). A rating of "insufficient information" was only given for eight ratings in total (involving only six cases).

To ensure that data met the assumptions for the statistical tests described above, all data were screened for outliers. Skewness and kurtosis were examined using the alpha value of .001 (e.g., $z > 3.29$), which is acceptable for small or moderate sized samples (Tabachnick & Fidell, 2001). Basic descriptive data is provided in Table 1 for the four risk assessment instruments. No transformations were necessary for the total or subtotal scores of these instruments. Moreover, the frequency distributions for the total scores of the four risk assessment instruments can be found in Appendices A through D.

Table 1: Basic Descriptive Statistics for Risk Assessment Instruments.

	Min.	Max.	M	SD	Skewness*	Kurtosis**
LSI-R total	0	48	28.29	10.920	-.425	-.695
LS/CMI total score	6	39	22.51	9.143	-.075	-.983
VRS: Static subtotal (prorated)	.00	16.00	7.483	3.952	.345	-.888
VRS: Dynamic subtotal (prorated)	1.00	48.42	17.888	11.883	.550	-.291
VRS Grand total (prorated)	3.00	63.00	25.247	14.913	.467	-.492
SIRR1 total score	-17	19	5.12	8.173	-.447	-.482

*Skewness Standard Error is .240; **Kurtosis Standard Error is .476

For the gender informed variables basic descriptive data is presented in Tables 2 and 3. Plots of the non-dichotomous variables were also explored to examine outliers (these have not been replicated here for reasons of brevity). As noted below, one individual had 100 episodes of

attempted suicide or self harm. When this outlier was removed, the variable did not provide much more information than its related measure of the presence or absence of such behaviour as all but 17 cases had none or one episode reported. Thus, this frequency variable was not included in additional analyses. Similarly, the number of children an offender had (in total or in her care prior to incarceration) was collected for descriptive purposes and thus was not transformed for normality. Subsequent analyses examined the presence or absence of childcare responsibility, not number of dependents.

Table 2: Frequencies for Dichotomous Gender Informed Variables.

	N	Percentage Present
Spouse or Common Law Partner at time of Index	101	52.5
Childcare Responsibility	101	67.3
Illegal Financial Support	101	89.1
History of Social Assistance	97	80.4
Self harm/suicide attempts	101	41.6
Substance Abuse	101	72.3
Childhood Physical Abuse	101	46.5
Childhood Sexual Abuse	101	49.5
Childhood Emotional Abuse	101	58.4
Adulthood Physical Abuse	100	68
Adulthood Sexual Abuse	101	21.8
Adulthood Emotional Abuse	100	59

Table 3: Basic Descriptive Statistics for Gender Informed Variables.

					Skewness		Kurtosis	
	Min.	Max.	M	SD	Statistic	Std. Error	Statistic	Std. Error
Number of children/ dependents residing with her when not incarcerated	0	4	.59	.982	1.802	.240	2.653	.476
Number of children/ dependents in total	0	11	1.86	2.117	1.741	.240	3.921	.476
Number of known self harm/suicide incidents*	0	100	1.92	10.227	9.308	.244	89.721	.483
Substance abuse need intensity **	0	2	1.30	.889	-.625	.240	-1.453	.476

N=101, unless otherwise indicated; *(N=98); **(0 = none,1 = some, 2 = considerable needs)

Data about the female offenders' abuse history was collected in great detail, including information about both the frequency of abuse and perpetrator type through childhood and adulthood. As reflected below offenders may have had multiple perpetrator types over time. Tables 4 and 5 outline this information. Of note, most female offenders who reported abuse histories were abused by family members, and there was also very little variability in the frequency of abuse; namely most abused female offenders were abused repeatedly over long periods of time. Given the limited variability in this data, subsequent analyses focused on the presence and absence of abuse. Thus, no transformations were made on this data.

Table 4: Percentage of Female Offenders Abused in Childhood by Frequency of Abuse and Perpetrator Type.

		Perpetrator Type			
	Frequency	Familial	Nonfamilial	Stranger	Unspecified
Physical Abuse					
	No Abuse	57.4	91.1	100	98
	A specific incident	1	2	0	2
	Several specific incidents	2	0	0	0
	Enduring/repeated for weeks/months/ years	39.6	6.9	0	0
Sexual Abuse					
	No Abuse	64.4	85.1	95	95
	A specific incident	6.9	5.9	4	2
	Several specific incidents	5	2	1	0
	Enduring/repeated for weeks/months/ years	23.8	6.9	0	3
Emotional Abuse					
	No Abuse	47.5	92.1	100	98
	A specific incident	1	0	0	1
	Several specific incidents	0	1	0	0
	Enduring/repeated for weeks/months/ years	51.5	6.9	0	1

Note some values do not add to 100% due to rounding.

Table 5: Percentage of Female Offenders Abused in Adulthood by Frequency of Abuse and Perpetrator Type.

		Perpetrator type			
	Frequency	Familial	Nonfamilial	Stranger	Unspecified
Physical Abuse					
	No Abuse	32.3	97	95	98
	A specific incident	5.1	2	4	2
	Several specific incidents	6.1	0	1	0
	Enduring/repeated for weeks/months/ years	56.6	1	0	0
Sexual Abuse					
	No Abuse	89	97	96	94
	A specific incident	1	3	2	2
	Several specific incidents	2	0	2	2
	Enduring/repeated for weeks/months/ years	8	0	0	2
Emotional Abuse					
	No Abuse	41	98	98	100
	A specific incident	1	0	1	0
	Several specific incidents	0	1	1	0
	Enduring/repeated for weeks/months/ years	58	1	0	0

Note some values do not add to 100% due to rounding.

As history of abuse may be especially relevant to recidivism when substance abuse is also present, Table 6 provides the frequencies for each type of abuse with regard to the presence of substance abuse. As noted above, approximately 72% of women had substance abuse concerns. Moreover, most abuse victims had substance abuse issues. However, substance abuse was also present in about half of the women who denied being abused, except when looking at adult sexual abuse. As adult sexual abuse had a low reported incidence, women who denied this form of abuse may have reported other forms of abuse.

Table 6: Frequency of each Abuse Type by Substance Abuse.

Abuse Type		Substance Abuse	
		Absent	Present
Childhood Physical Abuse	Absent	20	34
	Present	8	39
Childhood Sexual Abuse	Absent	23	28
	Present	5	45
Childhood Emotional Abuse	Absent	19	23
	Present	9	50
Adulthood Physical Abuse	Absent	17	15
(N=100)	Present	11	57
Adulthood Sexual Abuse	Absent	26	53
	Present	2	20
Adulthood Emotional Abuse	Absent	20	21
(N=100)	Present	8	51

N= 101, unless otherwise indicated

Outcome data were also explored to determine which variables required transformation. The untransformed descriptive statistics are found in Tables 7 and 8.

Table 7: Descriptive Statistics for Institutional Misconduct and Revocation Variables.

	Min.	Max.	M	SD	Skewness*	Kurtosis**
No. of Misconduct Charges during Index	0	52	3.98	7.796	3.645	16.648
No. of Minor Misconduct Charges during Index	0	28	2.72	4.928	2.804	8.896
No. of Serious Misconduct Charges during Index	0	24	1.26	3.478	4.759	25.365
No. of Misconduct Charges After Assessment Only	0	18	1.71	3.407	2.872	8.793
No. of Minor Misconduct After Assessment Only	0	12	1.16	2.284	2.752	8.264
No. of Serious Misconduct After Assessment Only	0	13	.55	1.622	5.334	35.926
No. of Revocations during Index	0	3	.64	.782	.981	.180

No. = number; *Skewness Standard Error is .240; **Kurtosis Standard Error is .476

Table 8: Descriptive Statistics for Recidivism Variables.

	Min.	Max.	M	SD	Skewness*	Kurtosis**
No. of Reconvictions	0	39	3.37	6.506	3.219	12.374
No. of Nonviolent Reconvictions	0	38	3.12	6.116	3.213	12.684
No. of Violent Reconvictions	0	4	.16	.543	4.694	26.856
No. of Reconvictions after WED	0	39	3.11	6.361	3.386	13.717
No. of Nonviolent Reconvictions after WED	0	38	2.92	5.988	3.366	13.988
No. of Violent Reconvictions after WED	0	3	.11	.445	4.677	23.440
Maximum Time Sentenced for Reconvictions	.00	913.00	59.068	149.440	3.465	13.604
Sum of Aggregate Incarceration Sentences	.00	1345.00	103.964	264.341	3.143	9.789
Sum of Aggregate Probation Sentences	.00	2374.38	303.154	532.018	2.027	3.757
Sum of Aggregate Conditional Sentences	.00	547.88	23.808	85.724	4.176	18.591
Maximum of Severity Index score ***	.00	13.00	3.436	4.168		
Minimum of Offence Categories Ranked by Seriousness (OCRS) Score***	2.00	25.00	16.703	8.479		

No. = number; *Skewness Standard Error is .240; **Kurtosis Standard Error is .476 ; ***Ordinal data, thus normality is not expected

As is evident above, a number of variables required transformations to improve normality. This was achieved by performing a series of transformations starting with a square root transformation, and proceeding to a log or inverse transformation if normality was not already achieved (as suggested by Tabachnick & Fidell, 2001). This process led to a log transformation being applied to the number of misconduct charges during the index sentence, the number of revocations, and the sum of aggregate probation sentences. Inverse transformations were used for the number of misconduct charges after the assessment period, the number of

minor misconduct changes (during the index sentence and following the assessment only), the number of reconvictions, the number of nonviolent reconvictions, maximum time sentenced for reconvictions, and the sum of aggregate incarceration sentences.

If no transformation was helpful in producing a sufficiently normal distribution, Tabachnick and Fidell (2001) suggest dichotomizing the variable. This was required for the number of violent reconvictions and number of serious misconduct charges (during the index sentence and following the assessment only). As well, the sum of aggregate conditional sentences could not be normalized because only 10 female offenders received these sentences for their reoffences and thus this variable was removed from subsequent analyses.

Note, the necessary data transformations were performed on non-normal data, but only statistics on the untransformed data are reported in this document unless differences exist. Further, basic descriptive statistics were provided on post WED reconvictions for information purposes only. The required transformations would have been identical to those required for the total reconviction data, but subsequent analyses focus only on the broad definition of reconvictions following the assessment or first release of the offender.

Time to reconviction data did not require transformation. Table 9 provides the descriptive statistics for these time data. As noted above these data were combined with follow up time data for nonrecidivists for subsequent survival analyses.

Table 9: Descriptive Statistics for Time (in Days) to Reconviction (Recidivists Only).

						Skewness		Kurtosis	
	N	Min.	Max.	M	SD	Statistic	Std. Error	Statistic	Std. Error
Time to 1st Reconviction	52	10.75	2789.00	1049.803	863.031	.628	.330	-.993	.650
Time to 1st Violent Reconviction	11	88.75	2476.75	985.386	1005.932	.679	.661	-1.712	1.279
Time to 1st Nonviolent Reconviction	50	10.75	2789.00	1082.140	864.036	.573	.337	-1.053	.662

3.1.2 Psychometric analyses.

See Tables 10 though 14 for interrater reliability for the risk assessment instruments. Additional item details are available in Appendices E through J. Generally reasonable levels of

reliability were achieved. As described in the methods section, attitudes and personal/emotional orientation domains were difficult to reliably rate as the information available on file was not always consistent between authors of different reports. The impact of this can be noted in reliability of the related domains on the LSI-R and LS/CMI. Further, information about a history of illegal financial support (unless adjudicated) was not always included in file information reliably (e.g., it may not be addressed routinely in the files, or it may be reported upon inconsistently by different authors). As this construct was unreliability or inconsistently reported, it is not surprising that reliability of its rating was also limited. Thus, interpretation of results based on these data should be made with caution.

Table 10: LSI-R Domain Analyses: Basic Descriptive Statistics and Interrater Reliability.

Domain	M	SD	Interrater Reliability (ICC)
LSI-R Total Score	28.29	10.92	.977
LSI-R Criminal History Domain	5.42	2.66	.949
LSI-R Education/Employment Domain	4.77	2.63	.968
LSI-R Finance Domain	1.08	0.60	.746
LSI-R Family/Marital Domain	2.11	1.16	.846
LSI-R Accommodation Domain	1.01	1.14	.963
LSI-R Leisure Domain	1.47	0.70	.706
LSI-R Associates Domain	3.18	1.22	.768
LSI-R Substance Abuse Domain	5.16	3.27	.953
LSI-R Personal/Emotional Domain	2.63	1.59	.546
LSI-R Attitudes Domain	1.47	1.45	.408

See Appendices E and F for LSI-R item reliability.

Table 11: LS/CMI Domain Analyses: Basic Descriptive Statistics and Interrater Reliability.

Domain	M	SD	Interrater Reliability (ICC)
LS/CMI Total	22.51	9.14	.940
Criminal History Domain	4.91	2.25	.993
Employment/Education Domain	4.46	2.65	.980
Family/Marital Domain	2.14	1.13	.846
Leisure Domain	1.26	0.73	.479
Associates Domain	2.97	1.04	.602
Substance Abuse Domain	3.52	2.77	.903
Attitudes Domain	1.50	1.46	.341
Antisocial Pattern Domain	1.75	1.35	.820

See Appendices G and H for LS/CMI item reliability.

Table 12: VRS Analyses: Basic Descriptive Statistics and Interrater Reliability.

Domain	M	SD	Interrater Reliability (ICC)
VRS Grand Total	25.247	14.913	.984
VRS Dynamic (prorated)	17.888	11.883	.862
VRS Static (prorated)	7.483	3.952	.981

See Appendix I for VRS item reliability.

Table 13: SIR-R1 Analyses: Basic Descriptive Statistics and Interrater Reliability.

Item	M	SD	Interrater Reliability (ICC)
SIR-R1 Total	5.12	8.17	.988

See Appendix J for SIR-R1 item reliability.

Table 14: Gender Informed Items Interrater Reliability.

Item	Interrater Reliability (ICC)
Spouse or Common Law Partner at time of Index	1.00
Childcare Responsibility (No. of children in total)	1.00
Illegal Financial Support	.382
History of Social Assistance	.763
Self harm/suicide attempts	1.00
Substance Abuse Issues	.873
Substance Abuse Needs Intensity	1.00
Childhood Physical Abuse	1.00
Childhood Sexual Abuse	1.00
Childhood Emotional Abuse	1.00
Adulthood Physical Abuse	1.00
Adulthood Sexual Abuse	.873
Adulthood Emotional Abuse	.573

No. = number

Internal consistency was examined using Cronbach's alpha (see Table 15). The three dynamic risk instruments had high alpha values, while the SIR-RI and the group of gender informed variables (which are not a developed scale) had lower values.

Table 15: Internal Consistency for the Risk Scales and the Gender Informed Variables.

	Internal Consistency of Instrument (Cronbach's alpha)
LSI-R	.908
LS/CMI	.909
VRS	.899
SIR-R1	.699
Gender Informed Variables*	.667

*Includes the 13 variables examined above

Table 16 provides the correlation values between the four risk assessment instruments. As expected the four risk instruments are highly correlated. The strongest relationship is seen between the total and dynamic risk subscale scores of the VRS, as expected given their shared items and that the latter composes more than two thirds of the total items in the scale. The strong relationship between the LSI-R and LS/CMI also underlines their shared content.

Table 16: Correlations Between Risk Instruments.

Risk Instrument	LSI-R Total	LS/CMI Total	VRS Static	VRS Dynamic	VRS Total	SIR-R1 Total
LSI-R Total	1	.880**	.643**	.705**	.734**	-.653**
LS/CMI Total		1	.648**	.675**	.711**	-.649**
VRS Static			1	.689**	.808**	-.637**
VRS Dynamic				1	.983**	-.508**
VRS Total					1	-.575**

**Correlation is significant at the 0.01 level (2-tailed).

Table 17 provides the correlation values between the 13 gender informed variables. As expected, Table 17 shows that substance abuse and childhood and adulthood abuse victimization are related. Further, while various abuse types share small to strong correlations, the relationship between physical and emotional abuse is especially strong in both adulthood and childhood experiences. Recall, that women were only identified as being emotionally abused if they provided specific examples or descriptions of such abuse to avoid redundancy when coding with other forms of abuse, as it is not unexpected for these experiences to co-occur. Self harm/suicidality is also related to both substance abuse and victimization experiences. Finally,

while childhood physical and sexual abuse and substance abuse were generally related to a history of social assistance, childhood emotional abuse was not, yet it was correlated with history of illegal financial support (a form of criminal behaviour). Moreover, history of social assistance and illegal financial support, both thought to be measuring economic pressures, are not correlated with each other suggesting that they are measuring distinct constructs and perhaps the latter more strongly reflects criminal behaviours generally than an assumed economic motivation underlying financially profitable crimes. However, recall that the low interrater reliability of illegal financial support may also be contributing to this difference.

Table 17: Correlation Between Gender Informed Variables.

	Presence of Children	Spouse/ Common law	Substance Abuse	Substance Abuse Intensity	Child Physical Abuse	Child Sexual Abuse	Child Emotional /Mental Abuse	Adult Physical Abuse	Adult Sexual Abuse	Adult Emotional /Mental Abuse	Self harm/ Suicide	Illegal Financial Support
Presence of Children	1											
Spouse/Common law	.182	1										
Substance Abuse	.087	.164	1									
Substance Abuse Intensity	.043	.230*	.883**	1								
Child Physical Abuse	.100	.172	.223*	.270**	1							
Child Sexual Abuse	.014	.070	.392**	.406**	.545**	1						
Child Emotional Abuse	.012	.243*	.330**	.351**	.666**	.474**	1					
Adult Physical Abuse (N= 100)	.248*	.199*	.384**	.418**	.289**	.286**	.285**	1				
Adult Sexual Abuse	.061	.166	.220*	.229*	.229*	.197*	.153	.301**	1			
Adult Emotional Abuse (N = 100)	.237*	.257**	.386**	.433**	.239*	.288**	.156	.605**	.330**	1		
Self harm/Suicide	.117	.200*	.343**	.307**	.341**	.450**	.345**	.354**	.090	.240*	1	
Illegal Financial Support	.163	.049	.138	.189	.135	.092	.221*	-.036	-.124	.097	.102	1
History of Social Assistance	.182	.050	.350**	.262**	.244*	.227*	.198	.202*	.017	.121	.176	.086

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 18 provides the correlations between the gender informed variables and gender informed composites. Being a single parent was only related to the two gender informed variables that combined to create the composite, but no other gender informed variables.

Substance abuse variables had large correlations with each composite, except the composite of substance abuse and adult sexual abuse which reflected only a moderate relationship. This may be due to the small incidence of reported sexual abuse in adulthood.

Child abuse variables were strongly correlated with the composites of substance abuse and all forms of child abuse, while adult abuse variables were mildly to moderately correlated with the child abuse composites. In general, adult abuse variables were also only moderately correlated to the composites of substance abuse and forms adult abuse, except those which they comprised (where a strong relationship was present, as expected). The child abuse variables also had moderate relationships with the composite of substance abuse and adult physical abuse, but smaller relationships with the composites including adult sexual or emotional abuse.

Self harm/suicidality had a moderate relationship with the composites of substance abuse and all forms of child abuse, as well as the composite of substance abuse and adult physical abuse. Suicidality/self harm was not related to the composite of substance abuse and adult sexual abuse, which again may be due to the incidence of the later. A small relationship was observed between the composite of substance abuse and adult emotional abuse and suicidality/self harm.

Finally, history of social assistance was mildly related to all of the abuse composites except those that included childhood emotional or adult sexual abuse. However, history of illegal financial support was not significantly related to any of the composites except substance abuse and child emotional abuse.

Table 18: Correlation Between Gender Informed Variables and Gender Informed Composites.

	Single Parent	Substance Abuse & Child Physical Abuse	Substance Abuse & Child Sexual Abuse	Substance Abuse & Child Emotional Abuse	Substance Abuse & Adult Physical Abuse (N= 100)	Substance Abuse & Adult Sexual Abuse	Substance Abuse & Adult Emotional Abuse (N= 100)
Spouse/ Common law	-.651**	.144	.055	.268**	.176	.125	.219*
Substance Abuse	-.061	.491**	.555**	.613**	.718**	.308**	.636**
Substance Abuse Intensity	-.158	.515**	.555**	.608**	.694**	.311**	.637**
Childcare Responsibility	.431**	.076	.072	.014	.121	.028	.205*
Illegal Financial Support	.146	.081	.122	.219*	.017	-.066	.103
History of Social Assistance (N = 97)	.082	.290**	.283**	.181	.291**	.058	.242*
Self harm/Suicide	-.119	.362**	.456**	.410**	.395**	.135	.248*
Child Physical Abuse	-.001	.850**	.522**	.545**	.315**	.284**	.262**
Child Sexual Abuse	-.038	.598**	.905**	.564**	.447**	.253*	.361**
Child Emotional Abuse	-.106	.587**	.514**	.835**	.366**	.218*	.260**
Adult Physical Abuse (N= 100)	-.097	.360**	.392**	.372**	.790**	.278**	.571**
Adult Sexual Abuse	-.059	.320**	.251*	.245*	.299**	.942**	.309**
Adult Emotional Abuse (N = 100)	-.024	.318**	.329**	.288**	.590**	.300**	.850**

N= 101 unless otherwise indicated; **. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

Table 19 examines the correlations between the four risk instruments and the gender informed variables. Not unexpectedly as most of the instruments include items addressing substance abuse, substance abuse variables were correlated with the risk instruments. However, the SIR-R1 does not include substance abuse in its items, thus its correlation with the substance abuse variables supports a more general relationship between substance abuse and measures of risk of criminal behaviour. Surprisingly, history of illegal financial support correlated significantly with the LS/CMI, but not the LSI-R which contains a specific financial domain, while both instruments should capture historic property offences equally. All instruments correlated with history of social assistance. Further, the history of abuse variables (with the exception of adulthood sexual abuse, and adulthood emotional abuse to some degree) also had a relationship with all of the risk instruments. Finally, self harm/suicidal behavior also largely had a moderate to strong relationship with the risk instruments.

Table 19: Correlations between Risk Instruments and Gender Informed Variables.

Gender Informed Variable	LSI-R Total	LS/CMI Total	VRS Static	VRS Dynamic	VRS Total	SIR-R1 Total
Childcare Responsibility	.071	.079	.027	.020	.027	.031
Spouse/Common law	.167	.060	.087	.176	.168	-.071
Presence of Substance Abuse	.706**	.623**	.493**	.543**	.563**	-.505**
Substance Abuse Intensity	.738**	.634**	.492**	.535**	.557**	-.485**
Child Physical Abuse	.368**	.322**	.528**	.310**	.381**	-.255**
Child Sexual Abuse	.377**	.321**	.471**	.335**	.389**	-.236*
Child Emotional Abuse	.436**	.355**	.528**	.340**	.407**	-.289**
Adult Physical Abuse (N= 100)	.352**	.308**	.347**	.326**	.354**	-.165
Adult Sexual Abuse	.191	.155	.105	.188	.174	-.081
Adult Emotional Abuse (N = 100)	.240*	.178	.173	.210*	.212*	-.036
Self harm/Suicide	.445**	.365**	.419**	.411**	.438**	-.230*
Illegal Financial Support	.194	.278**	.177	.096	.122	-.143
History of Social Assistance (N=97)	.264**	.235*	.232*	.211*	.225*	-.204*

N = 101, unless otherwise indicated; No. = number; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

The correlations between the outcome measures are provided in Appendix K.

3.1.3. Predictive validity.

3.1.3.1 Basic predictive validity.

Table 20 outlines the predictive validity of the four risk assessment instruments and institutional misconduct, revocations, recidivism, and reoffence severity. Overall, the

correlations above reflect a consistent relationship between the risk assessment instruments and institutional misconduct, revocation, recidivism, and severity of reoffence, using of number of different definitions. Only time to first violent reconviction was not significantly related to any of the risk instrument scores; however, this may have been due to there being only 11 cases of violent recidivism in this study. Further, the SIR-R1 scale did not correlate with time to first reconviction, for general or violent recidivism (although it did correlate with time to non-violent recidivism only). Most of the correlations reflect a moderate sized relationship, with some of the dichotomous variables reflecting large correlations at times. Number of revocations also had a large relationship with the LS/CMI total score.

Table 20: Correlations between Risk Instruments and Outcome.

	LSI-R Total	LS/CMI Total	VRS Total	SIR-R1 Total
Presence of Misconduct during Index	.404**	.502**	.454**	-.287**
No. of Misconduct Charges during Index	.388**	.453**	.530**	-.285**
No. of Minor Misconduct Charges during Index	.410**	.490**	.525**	-.319**
Presence of Serious Misconduct Charges during Index	.366**	.406**	.441**	-.245*
Presence of Misconduct After Assessment Only	.464**	.502**	.510**	-.390**
No. of Misconduct Charges After Assessment Only	.307**	.331**	.413**	-.250*
No. of Minor Misconduct After Assessment Only	.341**	.390**	.423**	-.339**
Presence of Serious Misconduct Charges After Assessment Only	.247*	.247*	.318**	-.196*
Presence of Revocations	.550**	.589**	.420**	.511**
No. of Revocations	.498**	.529**	.396**	-.475**
Recidivism Present	.419**	.471**	.478**	-.544**
No. Reconvictions	.282**	.306**	.302**	-.401**
Nonviolent Reconviction Present	.382**	.440**	.466**	-.545**
No. Nonviolent Reconvictions	.268**	.291**	.286**	-.399**
Violent Reconvictions Present	.289**	.295**	.390**	-.267**
Time to 1st Reconviction (n = 52)	-.372**	-.345*	-.349*	.260
Time to 1st Nonviolent Reconviction (n = 50)	-.366**	-.355*	-.351*	.284*
Time to 1st Violent Reconviction (n = 11)	.071	.089	-.212	-.351
Maximum Time Sentenced for Reconvictions	.323**	.347**	.283**	-.397**
Sum of Aggregate Incarceration Sentences	.340**	.369**	.309**	-.435**
Sum of Aggregate Probation Sentences	.226*	.267**	.309**	-.307**
Maximum of Severity Index score ***	.407**	.458**	.508**	-.575**
Minimum of Offence Categories Ranked by Seriousness (OCRS) Score***	-.386**	-.432**	-.469**	.568**

N=101, except where indicated; No. = number; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed). *** Spearman correlations for ranked data, remaining are Pearson correlations.

3.1.3.2 Relative/receiver operating characteristics (ROC).

Consistent with the correlations above, regardless of base rates of the outcomes, the ROC calculations continue to reflect large AUCs for most of the risk instruments and outcomes (greater than .71, Rice and Harris, 1995; see Table 21). The SIR-RI scale had only moderate sized AUCs with institutional misconduct throughout the index sentence, and all but the VRS had moderate AUCs for serious institutional misconduct after the assessment period. All relationships with recidivism and revocations were large. Of note, none of the confidence intervals were at or below .50 indicating that all of the predictions were consistently above chance. Further, all of the confidence intervals of the risk assessment instruments overlapped, suggesting they do not significantly differ from one another in their predictive validity (at least on these dichotomous outcome variables when unaffected by base rates, but not accounting for shared variance between the instruments).

Table 21: Areas Under the Curve and Confidence Intervals for Risk Instruments

Outcome	LSI-R		LS/CMI		VRS		SIR-R1	
	AUC	CI	AUC	CI	AUC	CI	AUC	CI
Institutional Misconduct during Index Offence	.747**	.650, .843	.798**	.711, .886	.778**	.688, .868	.678**	.573, .783
Serious Institutional Misconduct during Index Offence	.716**	.617, .814	.749**	.654, .843	.774**	.675, .873	.666**	.556, .776
Institutional Misconduct After Assessment	.778**	.685, .870	.797**	.708, .887	.809**	.724, .895	.734**	.631, .837
Serious Institutional Misconduct After Assessment	.658*	.553, .762	.670*	.561, .779	.763**	.617, .855	.650*	.525, .776
Revocation	.814**	.730, .898	.842**	.764, .920	.757**	.663, .850	.805**	.721, .888
Recidivism	.738**	.640, .837	.770**	.675, .865	.778**	.687, .869	.833**	.754, .913
Nonviolent Recidivism	.717**	.616, .818	.753**	.655, .850	.769**	.677, .861	.832**	.752, .911
Violent Recidivism	.783**	.664, .902	.772**	.636, .909	.835**	.737, .934	.769**	.648, .889

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed)

3.1.3.3 Additional relative predictive validity.

To determine if the risk instruments performed equally in predicting the outcome measures Meng, Rosenthal, and Rubin's (1992) method to test the significance of the difference between the correlations with a common variable in the same sample was employed (i.e., r(y, x1) = r(y, x2), where y = outcome, x1 = first predictor, x2 = second predictor). This was facilitated by use of a macro provided by Andrew Hayes (retrieved June 23, 2010 from http://www.comm.ohio-state.edu/ahayes/). Note, all scores on the SIR-R1 scale needed to be reflected before comparison as the original instrument predicts success not recidivism like the other instruments (thus the correlations with the SIR-R1 below are of the same magnitude as those reported above, but no longer negative unless the outcome was also transformed). See Table 22 and 23 for these comparisons. Note, only comparisons with at least one significant difference are listed. All other comparisons reflected no differences between the instruments.

The LS/CMI outperformed the LSI-R and the SIR-R1, but not the VRS, by sharing a stronger relationship with the presence of institutional misconduct during the index sentence. The VRS also had a significantly larger relationship with the presence of institutional misconduct during the index offence than the SIR-R1.

The number of institutional misconduct incidents was more strongly correlated with the VRS than all of the other instruments except the LS/CMI. The LS/CMI was also more strongly correlated with the number of institutional misconduct incidents than the SIR-R1 only. Note, when examining the log transformed data, the LS/CMI also outperformed the LSI-R ((r (y, x1) = .4865, r (y, x2) = .5813, z = -2.2949, p=.0217), but the VRS no longer had a significantly stronger relationship with the number of institutional misconduct incidents than the LSI-R (r (y, x1) = .4865, r (y, x2) = .5994, z = -1.8796, p=.0602). The remaining results were unchanged. This suggests that the differences between the performance of LS/CMI and LSI-R more likely reflect a small statistical difference than a clinically relevant differences in variance explained.

The number of minor institutional misconduct incidents had a stronger correlation with the LS/CMI and the VRS than with the SIR-R1 scale. The two instruments did not outperform each other or the LSI-R. However, once again when looking at the inverse transformed data, the LS/CMI also outperformed the LSI-R ((r (y, x1) = -.4728, r (y, x2) = -.5808, z = 2.6013, p=.0093), while no other differences from the untransformed data were observed.

Finally, the number of serious institutional misconduct incidents had a significantly stronger correlation with the LS/CMI and the VRS than the SIR-R1 scale.

When looking at institutional misconduct that occurred following the assessment period only (that is, not including any misconduct that occurred during the file review period), no significant differences in the performance of the assessment instruments were found with the untransformed data. When the inverse transformation of the total number of incidents of institutional misconduct after the assessment period only was examined the VRS did outperform the SIR-R1 scale (((r (y, x1) = -.5539, r (y, x2) = -.3944, z = -2.0146, p = .0439)), but no other significant differences were noted. Further, of note, this comparison on the untransformed data did reflect a trend toward the VRS outperforming the SIR-R1 scale ((r (y, x1) = .4134, r (y, x2) = .2501, z = 1.8851, p = .0594).

Table 22: Comparing Nonindependent Correlations for Misconduct Incidents during the Index Sentence.

Outcome		LSI-R vs. LS/CMI	LSI-R vs. VRS	LSI-R vs. SIR-R1	LS/CMI vs. VRS	LSCMI vs. SIR-R1	VRS vs. SIR-R1
Presence of Misconduct during Index	r (y, x1)	.4036	.4036	.4036	.5019	.5019	.4536
	r (y, x2)	.5019	.4536	.2867	.4536	.2867	.2867
	Z	-2.2457*	-.7624	1.4990	.7304	2.8203**	1.9663*
No. of Misconduct Charges during Index	r (y, x1)	.3882	.3882	.3882	.4527	.4527	.5301
	r (y, x2)	.4527	.5301	.2846	.5301	.2846	.2846
	Z	-1.4478	-2.2099*	1.3217	-1.1837	2.1699*	2.9623**
No. of Minor Misconduct Charges during Index	r (y, x1)	.4098	.4098	.4098	.4897	.4897	.5255
	r (y, x2)	.4897	.5255	.3187	.5255	.3187	.3187
	Z	-1.8237	-1.8118	1.1783	-.5542	2.2526*	2.5170*
Presence of Serious Misconduct Charges during Index	r (y, x1)	.3659	.3659	.3659	.4059	.4059	.4409
	r (y, x2)	.4059	.4409	.2455	.4409	.2455	.2455
	Z	-.8817	-1.1275	1.5174	-.5111	2.0284*	2.2712*

No. = number; y = outcome, x1 = first predictor, x2 = second predictor. ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

The VRS was outperformed by the LSI-R and the LS/CMI when predicting the presence of revocations. The LS/CMI also outperformed the VRS on the number of revocations, sharing a significantly stronger relationship.

The instruments did not differ in their predictive utility when examining the presence or number of incidents of reconvictions overall. However, there was one significant difference when examining the inverse transformed data on the number of reconvictions. Namely, the SIR-R1 better predicted the number of reconvictions than the LSI-R ((r (y, x1) = -.4027, r (y, x2) = -.5519, z=2.0766, p=.0378), but not the other instruments.

When examining the presence of nonviolent recidivism, the SIR-RI outperformed the LSI-R on the presence of nonviolent reconvictions, but none of the other instruments differed from each other. None of the instruments differed in their predictive validity with regard to number of nonviolent reconvictions either. However, when examining the inverse transformed

data for the number of nonviolent reconvictions, the SIR-R1 outperformed the LSI-R again ((r (y, x1) = -.3592, r (y, x2) = -.5354, z=2.4013, p=.0163), but not the other instruments. The instruments did not differ with regard to the presence of violent reconvictions.

Table 23: Comparing Nonindependent Correlations for Revocations and Recidivism.

Outcome		LSI-R vs. LS/CMI	LSI-R vs. VRS	LSI-R vs. SIR-R1	LS/CMI vs. VRS	LSCMI vs. SIR-R1	VRS vs. SIR-R1
Presence of Revocations	r (y, x1)	.5496	.5496	.5496	.5890	.5890	.4203
	r (y, x2)	.5890	.4203	.5112	.4203	.5112	.5112
	Z	-.9836	2.0516*	.5549	2.6184**	1.1473	-1.1385
No. of Revocations	r (y, x1)	.4980	.4980	.4980	.5294	.5294	.3965
	r (y, x2)	.5294	.3965	.4751	.3965	.4751	.4751
	Z	-.7466	1.5675	.3202	1.9939*	.7646	-.9635
Nonviolent Reconviction Present	r (y, x1)	.3821	.3821	.3821	.4403	.4403	.4662
	r (y, x2)	.4403	.4662	.5453	.4662	.5453	.5453
	Z	-1.2981	-1.2798	-2.2493*	-.3864	-1.4702	-1.0231

No. = number; y = outcome, x1 = first predictor, x2 = second predictor. ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

Looking at a number of measures of offence severity (i.e., time to first reconviction, time to first nonviolent reconviction, time to first violent reconviction, maximum time sentenced for a reconviction, sum of aggregate incarceration sentences, sum of aggregate probation sentences, and sum of conditional sentences), the instruments did not significantly differ from each other in the strength of their predictions. When examining the log transformed data for the sum of aggregate probation sentences, the LS/CMI did predict significantly better than the LSI-R ((r (y, x1) = .2814, r (y, x2) = .3838, z=-2.2027, p=.0276), but none of the other instruments differed from each other.

3.1.4 Predictive validity of risk levels.

Women with higher risk scores on each of the traditional needs/risk assessment instruments were predicted to have higher and faster rates of general and violent recidivism than those with lower scores, as illustrated by survival analysis. Note, the risk levels are based on suggestions in manuals where available; however, these suggestions are based on males samples for the SIR-R1. The risk levels for the VRS are based on creating quartiles to form three groups (top 25%, middle 50%, and top 25%), as in previous literature of on the VRS and related

instruments (e.g., VRS-SO, VRS-YV). See Appendix L for the specific general recidivism rates by risk level for each risk assessment instrument and Appendix M for the specific violent recidivism rates by risk level for each risk assessment instrument.

3.1.4.1 General recidivism.

Overall, the average time for recidivists to be first reconvicted was 1049.80 days (SD= 863.03), or just under three years, with the earliest first reconviction occurring only 10 days after assessment and release and the last occurring after 2789 days, or just over seven and a half years (near the end of the follow up period for most women).

Figure 2 provides the survival curves for the LSI-R. A chi square analysis showed a significant difference between the LSI-R risk levels on time to first reconviction (Wilcoxon (4) = 24.087, p <.001). Looking at the table of means below (Table 24) we can see that the confidence intervals for the low and low/moderate groups overlap, but neither overlaps with the moderate group. However, the moderate group does not differ from the medium/high and high groups. This suggests that the five risk levels do not differentiate as well as would be expected; however, this may be due to both the small sample size and the fact that the risk levels were based on male offender and provincial female offender data. A two level differentiation is clear, looking at the means below, and in Figure 1.

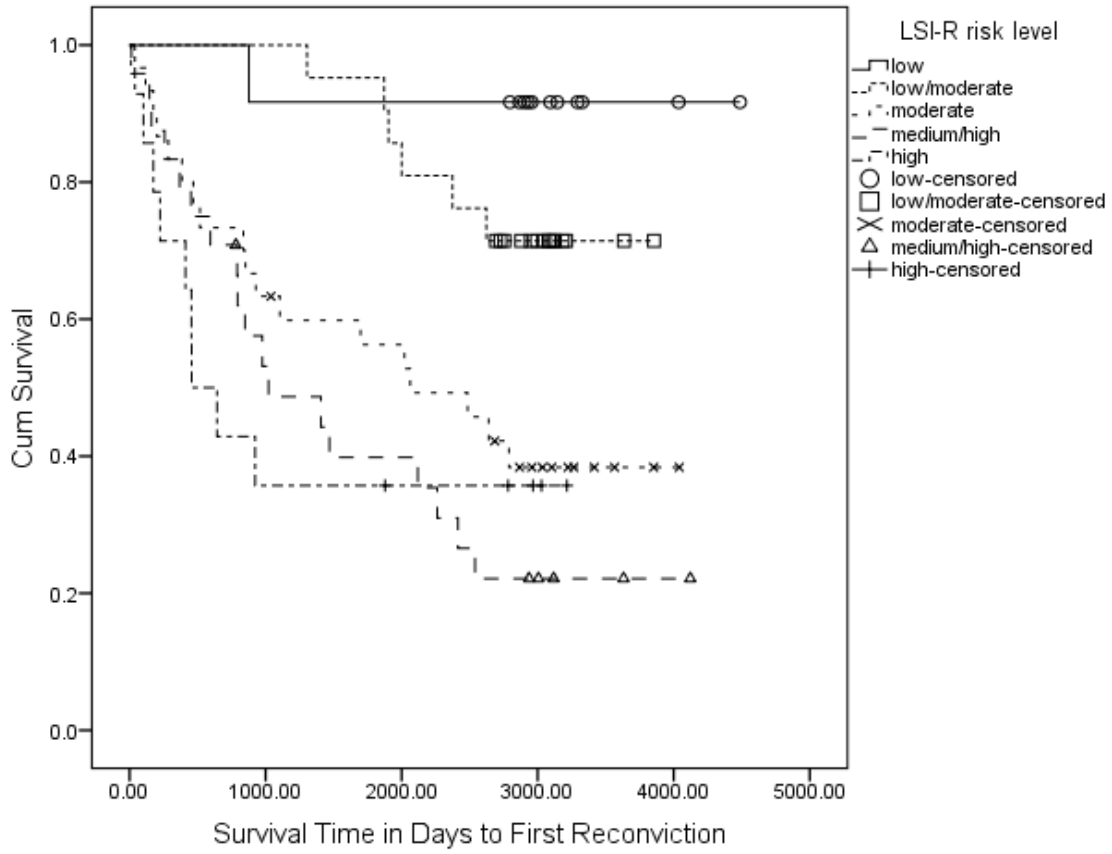

Figure 1: Survival Analysis: Cumulative general recidivism failure rate as a function of LSI-R Risk Levels.

Table 24: Mean Survival Time for General Recidivism for each LSI-R Risk Level.

LSI-R risk level	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
low	12	4186.146	288.046	3621.576	4750.716
low/moderate	21	3328.524	188.001	2960.043	3697.005
moderate	30	2240.010	293.934	1663.900	2816.120
medium/high	24	1731.766	302.710	1138.454	2325.078
high	14	1393.036	367.333	673.063	2113.008

a. Estimation is limited to the largest survival time if it is censored.

Figure 2 provides the survival curves for the LS/CMI. A chi square analysis showed a significant difference between the LS/CMI risk levels on time to first reconviction (Wilcoxon (3) = 31.747, p <.001). Looking at the table of means below (Table 25) we can see that the confidence intervals for the medium group differ from all of the others, while the high risk and very high risk group overlap slightly. Of note, there were no women who scored in the very low risk level and all those who scored in the low risk group did not recidivate. This suggests at least a three level differentiation is possible despite the small sample size.

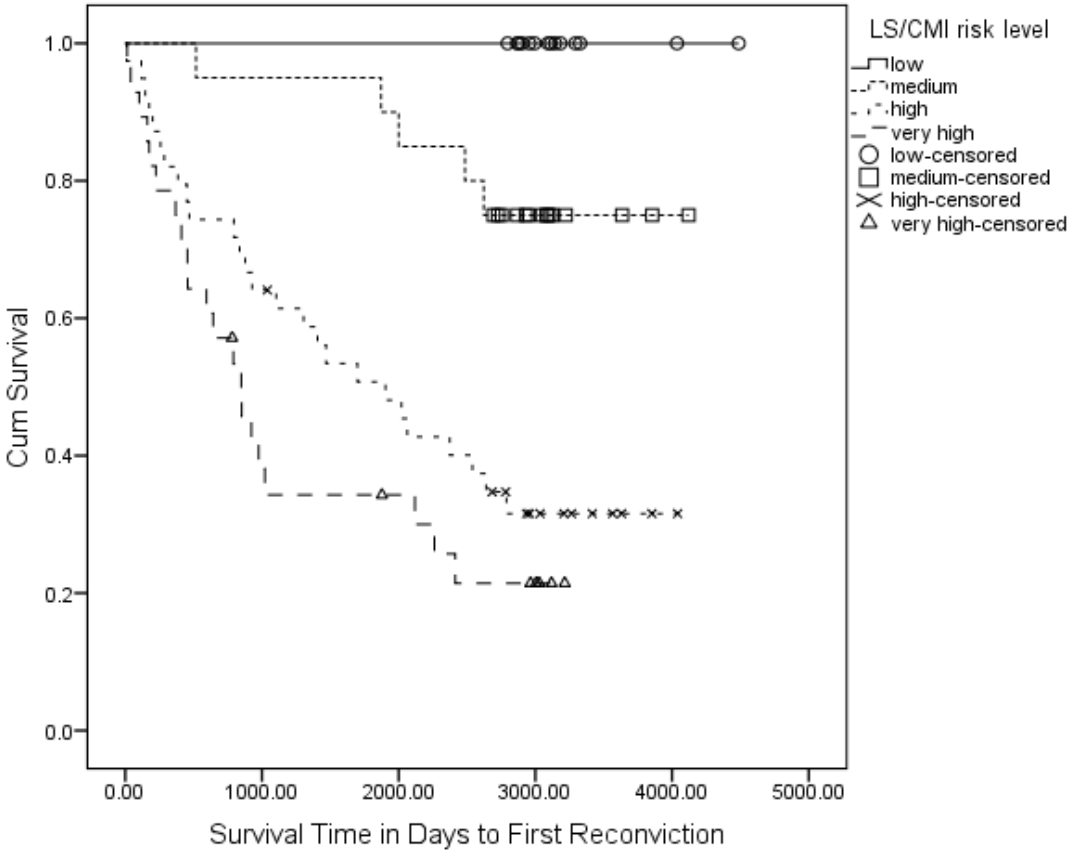

Figure 2: Survival Analysis: Cumulative General Recidivism Failure Rate as a Function of LS/CMI Risk Levels.

Table 25: Mean Survival Time for General Recidivism for each LS/CMI Risk Level.

LS/CMI Risk Level	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
medium	20	3565.612	230.726	3113.389	4017.836
high	39	2063.856	246.643	1580.436	2547.276
very high	28	1316.201	226.617	872.031	1760.371

N = 14 for the low risk level

a. Estimation is limited to the largest survival time if it is censored.

Figure 3 provides the survival curves for the VRS. A chi square analysis showed a significant difference between the VRS risk levels on time to first reconviction (Wilcoxon (2) = 26.727, p <.001). None of the three confidence intervals overlap, suggesting good differentiation between risk levels (see Table 26). While the three risk levels appear to fit the data better than the risk levels for the LSI-R and LS/CMI it is not surprising as the cut off scores follow the truncation method used in past research with men and youth, but are based on the frequency distribution of this data not predetermined cut off scores. Further, a three level differentiation will be more easily replicated on a smaller sample than a five level differentiation as prescribed by the LSI-R and LS/CMI.

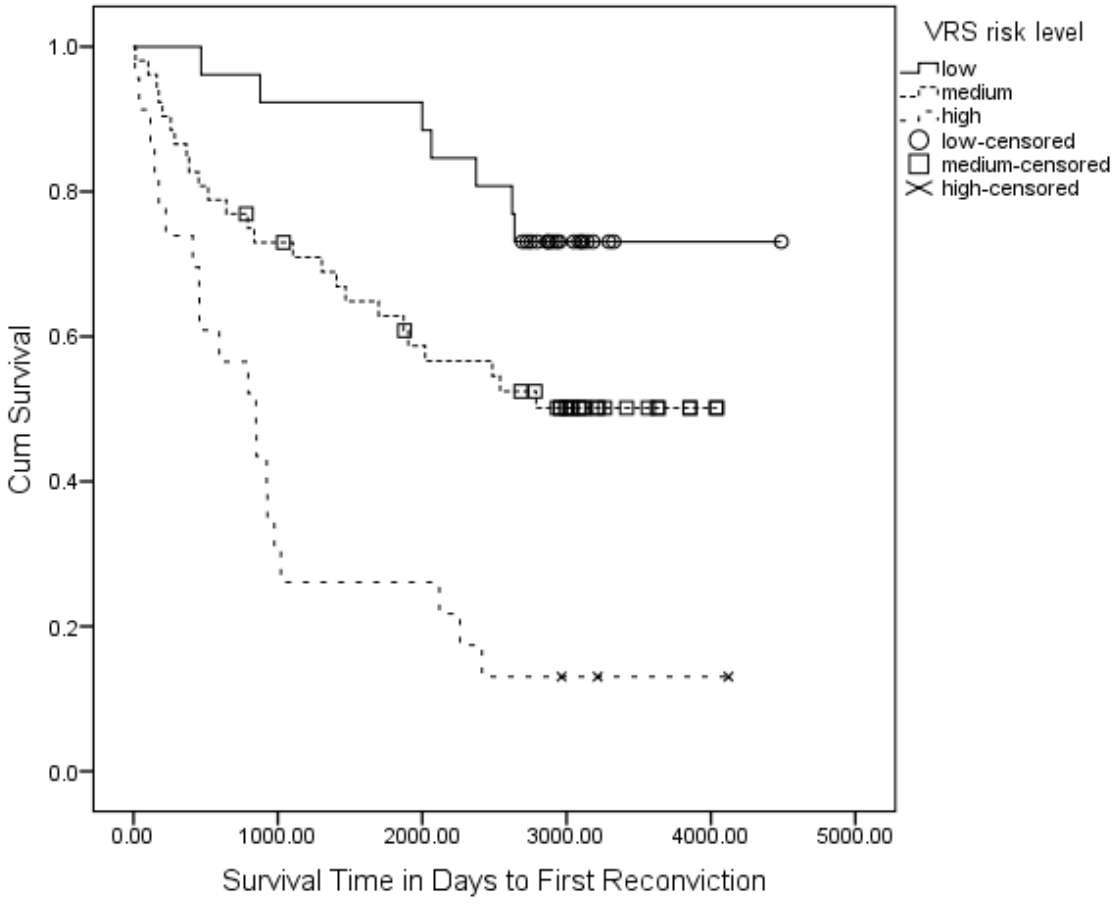

Figure 3: Survival Analysis: Cumulative General Recidivism Failure Rate as a Function of VRS Risk Levels.

Table 26: Mean Survival Time for General Recidivism for each VRS Risk Level.

VRS Risk Levels	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
low	26	3780.596	242.135	3306.012	4255.180
medium	52	2557.741	225.218	2116.313	2999.170
high	23	1222.815	270.901	691.850	1753.780

a. Estimation is limited to the largest survival time if it is censored

Figure 4 provides the survival curves for the SIR-R1. A chi square analysis showed a significant difference between the SIR-R1 risk levels on time to first reconviction (Wilcoxon (4) = 34.789, p <.001). Looking at Table 27, the low risk level does not overlap with the confidence interval of any other risk level. However, the low to moderate risk and moderate risk levels do not differ from one another. And although the low to moderate risk level does differ from the moderate to high risk level, the moderate risk level does not. The high risk level also overlaps with all but the low level. The utility of these risk levels appear to be the poorest of the four instruments; however, again a five risk level differentiation is difficult to achieve well with such a small sample. Further, while examining the possibility of better differentiation based on the frequency distribution of this data into thirds improved the performance of risk levels created for this instrument (low risk level with scores from 10 to 19, a moderate risk level group with scores from 2 to 9, and a high risk level group with scores from 1 to -17), the small sample size available here would likely limit its utility beyond this study. However, these cut off scores do improve the step-like progression from low to high risk. Chi square and survival analyses demonstrating the suitability of these revised risk levels are in Appendix N.

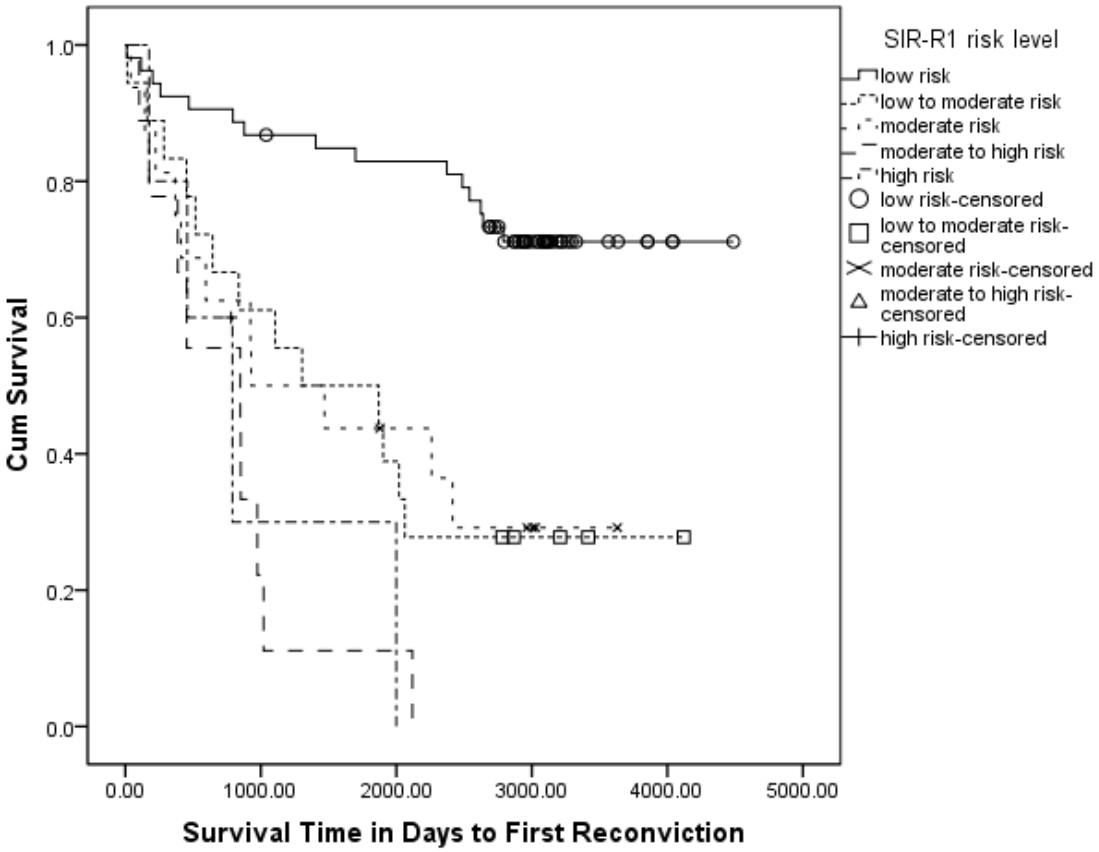

Figure 4: Survival Analysis: Cumulative General Recidivism Failure Rate as a Function of SIR-R1 Risk Levels.

Table 27: Mean Survival Time for General Recidivism for each SIR-R1 Risk Level.

SIR-R1 Risk Levels	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
low risk	53	3607.489	205.709	3204.300	4010.678
low to moderate risk	18	1876.778	358.275	1174.559	2578.997
moderate risk	16	1718.813	353.849	1025.269	2412.356
moderate to high risk	9	770.194	203.556	371.225	1169.164
high risk	5	963.425	405.432	168.779	1758.071

a. Estimation is limited to the largest survival time if it is censored

3.1.4.1 Violent recidivism.

The average time for the first violent reconviction for this sample was 985.39 days (SD= 1005.93), or over two and a half years (range = 88.75 to 2476.75 days). Recall only 11% of the total sample had a violent reconviction after the over seven year follow up period.

Figure 5 provides the survival curves for the LSI-R. A chi square analysis showed a significant difference between the LSI-R risk levels on time to first violent reconviction (Wilcoxon (4) = 12.629, p=.013). Table 28 shows that the confidence intervals for the moderate and medium/high risk levels overlap, while the high risk level is distinct. Of note, none of the women in the low or low/moderate risk level groups recidivated violently. This suggests a possible three level differentiation with these cut-off scores.

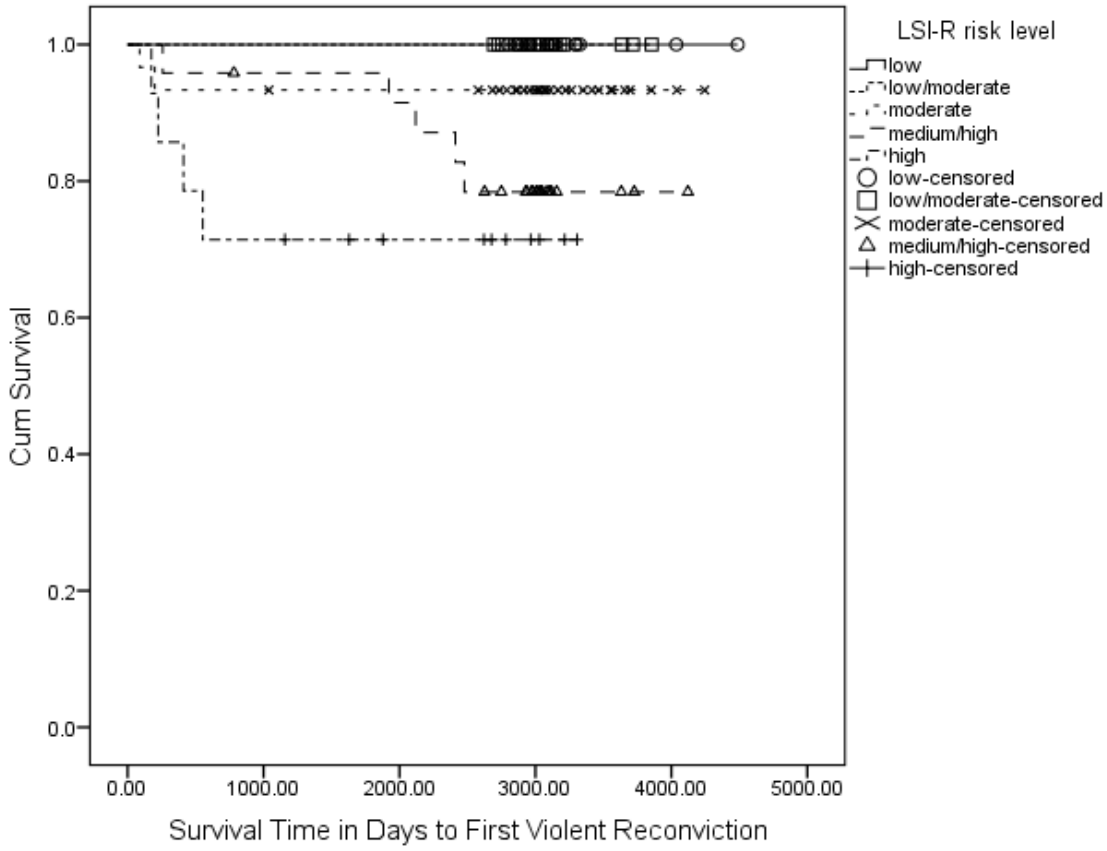

Figure 5: Survival Analysis: Cumulative Violent Recidivism Failure Rate as a Function of LSI-R Risk Levels.

Table 28: Mean Survival Time for Violent Recidivism for each LSI-R Risk Level.

LSI-R risk level	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
moderate	30	3971.658	186.747	3605.635	4337.682
medium/high	24	3630.940	207.702	3223.845	4038.035
high	14	2458.571	358.648	1755.620	3161.522

N = 12 for the low risk level; N = 21 for the low/moderate risk level

a. Estimation is limited to the largest survival time if it is censored

Figure 6 provides the survival curves for the LS/CMI. A chi square analysis showed a significant difference between the LS/CMI risk levels on time to first violent reconviction (Wilcoxon (3) = 10.126, p=.018). Table 29 shows that the confidence intervals for the high and very high group overlap; however, none of the women in the low or medium risk level groups had a violent reconviction, which explains the significant chi square analysis. As previously noted none of the women has scores in the very low risk level. Here a two level differentiation is suggested.

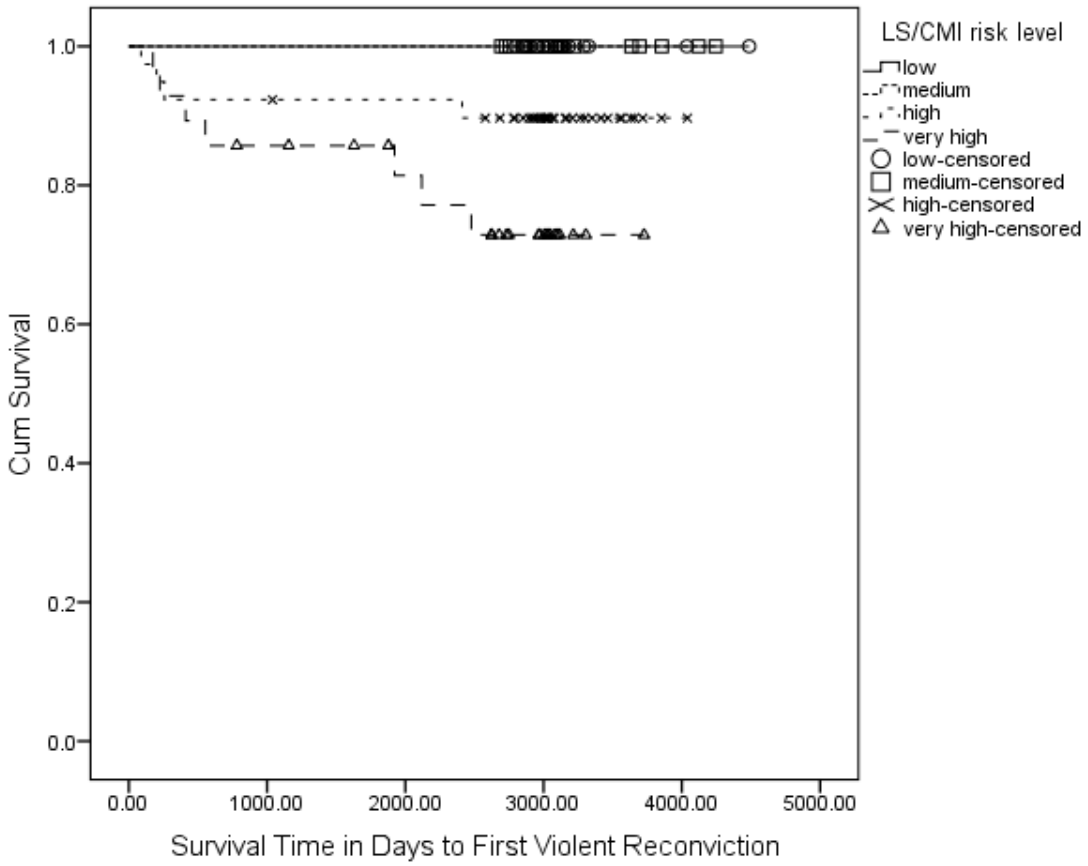

Figure 6: Survival Analysis: Cumulative Violent Recidivism Failure Rate as a Function of LS/CMI Risk Levels.

Table 29: Mean Survival Time for Violent Recidivism for each LS/CMI Risk Level.

LS/CMI Risk Level	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
high	39	3700.998	168.085	3371.551	4030.446
very high	28	3043.368	234.960	2582.847	3503.889

N = 14 for the low risk level; N = 20 for the medium risk level

a. Estimation is limited to the largest survival time if it is censored

Figure 7 provides the survival curves for the VRS. A chi square analysis showed a significant difference between the VRS risk levels on time to first violent reconviction (Wilcoxon (3) = 13.763, p=.001). Table 30 shows that the confidence intervals for the medium and high risk group overlap slightly; however, none of the women in the low risk level group had a violent reconviction, which explains the significant chi square analysis. These risk level groupings appear promising, given the limited sample size and base rate of violent reconviction.

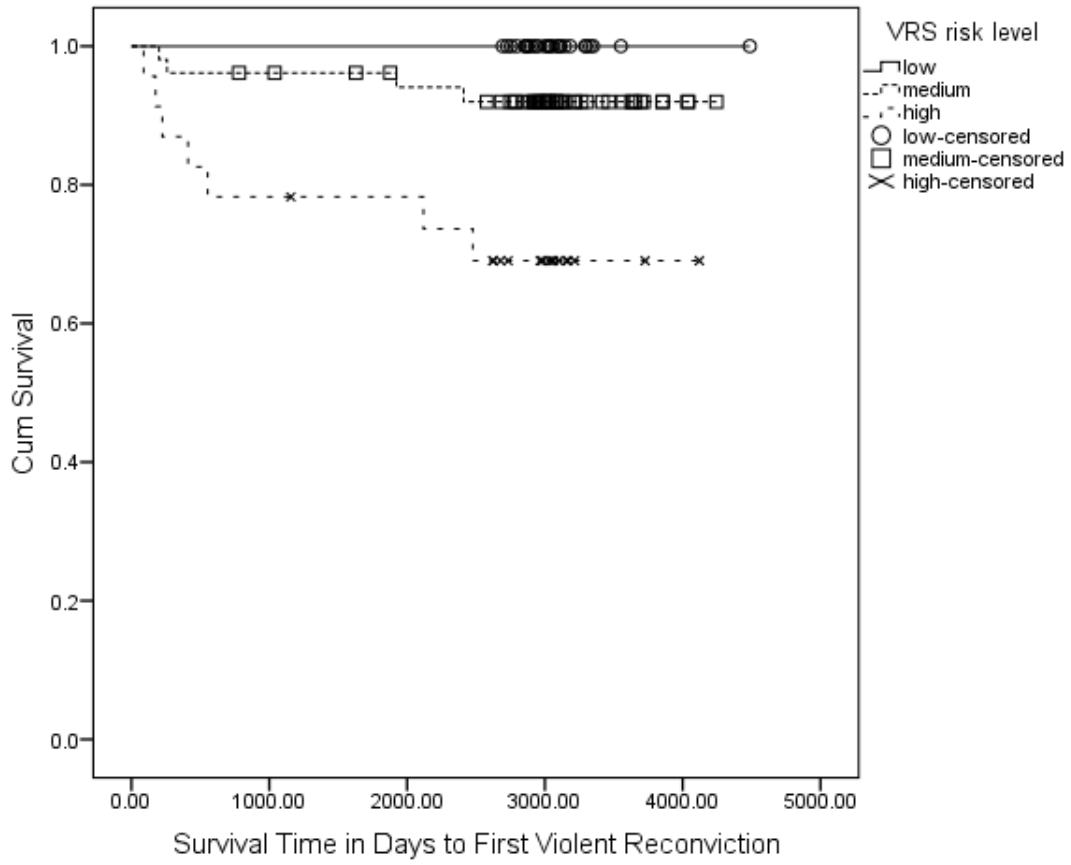

Figure 7: Survival Analysis: Cumulative Violent Recidivism Failure Rate as a Function of VRS Risk Levels.

Table 30: Mean Survival Time for Violent Recidivism for each VRS Risk Level.

VRS Risk Level	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
medium	52	4003.673	120.911	3766.688	4240.658
high	23	3120.359	331.014	2471.572	3769.146

N = 26 for the low risk level

a. Estimation is limited to the largest survival time if it is censored.

Figure 8 provides the survival curves for the SIR-R1. A chi square analysis showed a significant difference between the SIR-R1 risk levels on time to first violent reconviction (Wilcoxon (4) = 11.472, p=.022). Of note, the low to moderate risk level had no women with violent reconvictions. Table 31 shows that the confidence interval for low risk level is differentiated from the remaining intervals, but none of them significantly differ from each other. Again the risk levels of the SIR-R1 seem to have the least utility and do not follow a steplike progression (that is, an increase in the rate of recidivism as the risk level increases), but as noted above with such a small rate of violent reconvictions in this sample, the data are especially limited by the small sample size.

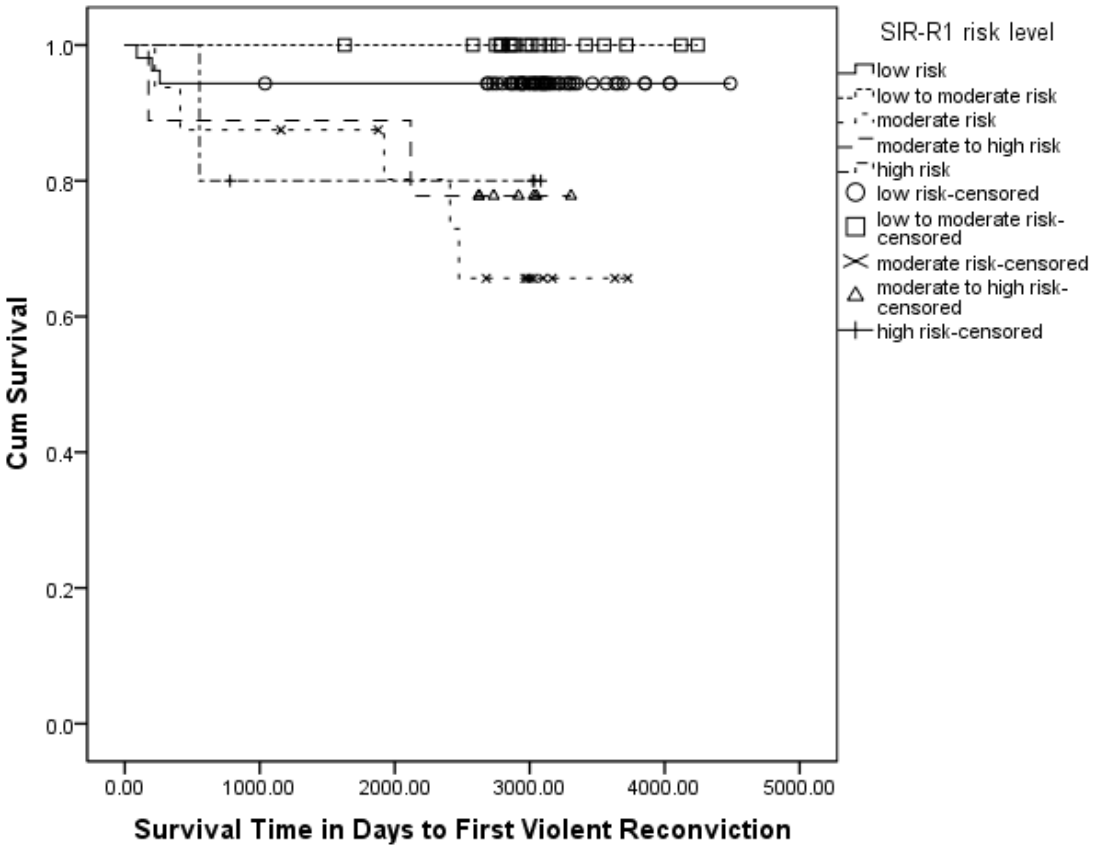

Figure 8 Survival Analysis: Cumulative Violent Recidivism Failure Rate as a Function of SIR-R1 Risk Levels.

Table 31: Mean Survival Time for Violent Recidivism for each SIR-R1 Risk Level.

SIR-R1 Risk Level	Mean[a] Survival Time				
				95% Confidence Interval	
	N	Estimate	Std. Error	Lower Bound	Upper Bound
low risk	53	4243.368	136.642	3975.550	4511.185
moderate risk	16	2981.979	299.772	2394.426	3569.532
moderate to high risk	9	2826.000	335.900	2167.636	3484.364
high risk	5	2577.150	452.446	1690.356	3463.944

N = 18 for the low to moderate risk level

a. Estimation is limited to the largest survival time if it is censored.

3.1.5. Incremental predictive validity.

3.1.5.1. Predictive validity of the gender informed variables.

Table 32 outlines the correlations between the gender informed variables and institutional misconduct and revocations. Given the number of variables being examined, consistency between multiple definitions was important to support to any possible findings.

Overall, few gender informed variables were correlated with revocations or institutional misconduct over numerous measures of the construct, other than substance abuse (which is tapped already by most of the traditional risk assessment instruments). Substance abuse had significant moderate correlations with revocation and significant small correlations with misconduct in general, except that the intensity of substance abuse was not related to the number of minor incidents of institutional misconduct until an inverse transform was applied to the latter variable (r=-.282, p=.004).

Adulthood sexual abuse had a small correlation with the presence of institutional misconduct during the index sentence. The inverse transform of the number of minor institutional misconduct incidents during the index offence also negatively correlated with adulthood sexual abuse (r=-.199, p=.046). Childhood emotional abuse was also mildly positively related to the number of minor misconduct incidents during the index offence, while an inverse transformation was required to reveal a relationship between childhood emotional abuse and number of minor incidents of institutional misconduct after the assessment period (r=-2.16, r=.030).

A history of illegal financial support and a history of social assistance support had small correlations with the presence and number of revocations. Childhood sexual abuse also had a small relationship with the presence, but not number, of revocations.

Table 32: Correlations between the Gender Informed Variables and Institutional Misconduct and Revocations.

Gender Informed Variables	Presence of Revocations	No. of Revocation	Presence of Misconduct during Index	No. of Misconduct during Index	No. of Minor Misconduct during Index	Presence of Serious Misconduct during Index	Presence of Misconduct After Assessment	No. of Misconduct After Assessment	No. of Minor Misconduct After Assessment	Presence of Serious Misconduct After Assessment
Childcare Responsibility	.071	.115	-.101	-.070	-.052	-.070	-.098	-.109	-.081	-.195
Spouse/Common law	.191	.150	.044	.092	.084	.009	.079	.060	.093	.022
Presence Substance Abuse	.412**	.398**	.259**	.212*	.218*	.232*	.343**	.235*	.218*	.273**
Substance Abuse Intensity	.442**	.413**	.227*	.216*	.202*	.228*	.330**	.207*	.159	.257**
Illegal Financial Support	.269**	.248*	.077	.147	.162	.101	.102	.120	.136	.030
History of Social Assistance (N=97)	.205*	.269**	.003	.058	.063	-.044	.075	.065	.078	-.030
Self harm/Suicide	.162	.180	.093	.137	.134	.073	.103	.024	.056	-.007
Child Physical Abuse	.106	.070	.019	.046	.065	-.123	.018	-.085	-.030	-.156
Child Sexual Abuse	.208*	.174	.031	.048	.096	.049	.089	-.062	-.008	-.043
Child Emotional Abuse	.159	.104	.069	.140	.198*	.056	.141	.059	.147	.007
Adult Physical Abuse (N= 100)	.174	.171	.083	.039	.036	-.050	.092	-.010	-.008	-.120
Adult Sexual Abuse	.074	.057	.222*	-.036	.015	.002	.139	.009	.048	.012
Adult Emotional Abuse (N = 100)	.092	.075	-.002	.049	.012	-.101	-.049	-.014	-.048	-.169

No. = number; N=101, except where indicated; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

Table 33 outlines the correlations between the gender informed variables and recidivism and severity of reoffence. Again, given the number of comparisons made, findings which were not replicated across a number of definitions of recidivism were interpreted with caution.

Again, few gender informed variables were correlated with recidivism over numerous measures of the construct, other than substance abuse which had significant small to moderate correlations with a variety of recidivism measures, but not violent reconvictions. History of illegal financial support also had small significant correlations with the presence of all recidivism and nonviolent reconvictions, as well as the inverse transform of the number of reconvictions (r= -.205, r= .039). This is not surprising as it likely reflects similar past behavior (e.g., instrumental nonviolent crimes). The presence of violent reconvictions did share a small significant correlation with self harm/suicide attempts as well as childhood emotional abuse.

Only two comparisons became significant when the transformed data was examined. Namely, the presence of self harm/suicide attempts had a small correlation with the inverse transform of the longest incarceration sentence earned during the follow up period (r=-.198, r=.047) suggesting those with less self harm behavior also earned shorter sentences. Also, adult sexual abuse had a small significant negative correlation with the log transform of the sum of aggregate probation sentences earned during the follow up period (r= -.201, p=.044). Of note, there was a trend (p=.052) for this correlation (r=-.194) when the raw data were examined.

The only other relationships found with the gender informed variables were lost when the transformed recidivism variables were examined. Specifically, the presence of a spouse had a small positive significant relationship with the number of nonviolent reconvictions, which was no longer present with the inverse transformation of the number of nonviolent reconvictions (r=-.169, p=.090, respectively). Finally, the presence of a spouse had a small significant correlation with the sum of the aggregate incarceration sentences earned during the follow up period, which did not exist when the inverse transform was examined (r=-.152, p=.128).

Table 33: Correlations between the Gender Informed Variables and Recidivism.

Gender Informed Variables	Recidivism Present	No. of Reconvictions	Nonviolent Reconviction Present	No. Nonviolent Reconvictions	Violent Reconvictions Present	Max of Time Sentenced	Sum of Aggregate Incarceration Sentences	Sum of Aggregate Probation Sentences
Childcare Responsibility	.084	.131	.056	.125	.040	.065	.042	.062
Spouse/Common law	.147	.189	.189	.201*	.078	.163	.203*	.101
Presence Substance Abuse	.328**	.223*	.304**	.227*	.146	.216*	.216*	.135
Substance Abuse Intensity	.326**	.228*	.317**	.232*	.134	.262**	.260**	.123
Illegal Financial Support	.233*	.098	.219*	.090	.122	.132	.127	.156
History of Social Assistance (N=97)	-.035	.128	-.050	.125	.059	.024	.072	.097
Self harm/Suicide	.096	.117	.049	.106	.221*	.094	.097	.064
Child Physical Abuse	-.008	-.050	.029	-.077	.184	.042	-.012	-.004
Child Sexual Abuse	-.029	.088	-.030	.075	.099	.061	.066	.130
Child Emotional Abuse	.146	.091	.152	.069	.230*	.052	.039	.142
Adult Physical Abuse (N= 100)	.014	.072	.029	.070	.104	.018	-.005	.004
Adult Sexual Abuse	-.064	-.100	-.043	-.097	-.107	.021	-.061	-.194
Adult Emotional Abuse (N = 100)	-.004	.130	.004	.134	.098	.051	.111	-.085

No. = number; N=101, except where indicated; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

When examining offence severity (see Table 34), once again substance abuse (presence and intensity) shares moderate significant correlations with these measures. Further, history of illegal financial support also shares a small significant relationship with both measures of severity. These findings are consistent with the other recidivism outcomes above.

Table 34: Spearman Correlations between the Gender Informed Variables and Offence Severity.

Gender Informed Variables	Minimum of Offence Categories Ranked by Seriousness (OCRS) Score	Maximum of Severity Index Score
Childcare Responsibility	-.126	.085
Spouse/Common law	-.166	.134
Presence Substance Abuse	-.310**	.322**
Substance Abuse Intensity	-.297**	.310**
Illegal Financial Support	-.199*	.198*
History of Social Assistance (N=97)	-.063	.027
Self harm/Suicide	-.075	.096
Child Physical Abuse	-.021	.063
Child Sexual Abuse	.001	.018
Child Emotional Abuse	-.128	.139
Adult Physical Abuse (N= 100)	-.030	.033
Adult Sexual Abuse	.104	-.137
Adult Emotional Abuse (N = 100)	.022	-.028

N=101, except where indicated; No. = number; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

Similarly, ROC analyses were conducted to examine these relationships with the gender informed variables without the influence of base rates on the outcome measures. Table 35 highlights the only variables where significant AUCs were found. All other gender informed variables yielded no significant AUC values for any other dichotomous outcome variable.

Of note, substance abuse was predictive of any and serious institutional misconduct after the assessment period, revocation, and any and nonviolent recidivism, but not violent recidivism. Childhood emotional abuse yielded the only AUC from the gender informed variables that was significant with regard to violent recidivism. That AUC reflected a moderate sized effect.

Table 35: Areas Under the Curve and Confidence Intervals for Gender Informed Variables with Outcome.

Outcome	Substance Abuse		Child Emotional Abuse	
	AUC	CI	AUC	CI
Institutional Misconduct After Assessment	.643*	.534, .752	.562	.446, .678
Serious Institutional Misconduct After Assessment	.641*	.521, .760	.496	.356, .637
Revocation	.688**	.582, .793	.582	.468, .697
Recidivism	.632*	.520, .744	.565	.450, .679
Nonviolent Recidivism	.621*	.509, .733	.568	.453, .682
Violent Recidivism	.600	.440, .760	.687*	.545, .829

N=97, ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

3.1.5.2. Predictive validity of the gender informed composites.

In addition to exploring the gender informed variables individually, past research has suggested some composites of these variables which may be uniquely valuable when predicting outcome. Namely, looking at being a single parent (Bonta et al., 1995), and past experiences of victimization in combination with substance abuse (McClellan et al., 1997, Rettinger, 1998) may be related to outcome regardless of whether each variable of the composite was related on its own. Table 36 shows the correlations of each of these composites with the revocation and institutional misconduct.

Being a single parent did not significantly correlate with any of the revocation or institutional misconduct variables, nor did the combination of substance abuse and childhood physical abuse. Substance abuse and adulthood physical abuse combined to create a moderate significant relationship with both measures of revocation, and a small significant relationship with the presence of institutional misconduct during the index offence. Additional relationships appeared when the transformed data was examined. Specifically, small significant relationships

between the combination of substance abuse and adulthood physical abuse and the log transform of the number of any misconduct incidents during the index sentence (r=.215, p=.032), the inverse transform of the number of minor misconduct incidents during the index sentence (r=-.285, p=.004), and the inverse transform of the number of minor misconduct incidents after the assessment only (r=-.230, p=.022) were revealed.

The combination of substance abuse and childhood sexual abuse had small significant correlations with both measures of revocation, as well as with the inverse transform of the number of minor institutional misconduct episodes. The combination of substance abuse and adulthood sexual abuse also had a small correlation with the inverse transform of the number of minor institutional misconduct episodes, but not with any other measures of revocation or institutional misconduct.

Perhaps the most interesting and consistent findings were for the combination of substance abuse and childhood emotional abuse. It had small significant correlations with the presence of revocations, the number of minor misconduct incidents during the index sentence, and the presence of misconduct after the assessment period. Further, when examining the transformed data three more small significant correlations appear; with the log transform of the number of misconduct incidents during the index sentence (r=.216, p=.030), with the inverse transform of the number of misconduct incidents after the assessment (r=-.227, p=.022) and with the inverse transform of the number of minor misconduct incidents after the assessment (r=-.264, p=.008). The composite of substance abuse and adulthood emotional abuse was only significantly related to the presence of revocations to a small degree.

Table 36: Correlations between the Gender Informed Composites and Revocations and Institutional Misconduct.

Gender Informed Composites	Presence of Revocations	No. of Revocations	Presence of Misconduct during Index	No. of Misconduct during Index	No. of Minor Misconduct during Index	Presence of Serious Misconduct during Index	Presence of Misconduct After Assessment	No. of Misconduct After Assessment	No. of Minor Misconduct After Assessment	Presence of Serious Misconduct After Assessment
Single Parent	-.102	-.086	-.125	-.061	-.064	-.041	-.074	-.032	-.072	-.112
Substance Abuse & Child Physical Abuse	.141	.128	.082	.068	.074	-.103	.074	-.065	-.019	-.123
Substance Abuse & Child Sexual Abuse	.224*	.206*	.105	.056	.091	.075	.133	-.065	-.019	-.039
Substance Abuse & Child Emotional Abuse	.208*	.174	.151	.168	.213*	.092	.209*	.090	.166	.053
Substance Abuse & Adult Physical Abuse (N=100)	.332**	.317**	.207*	.142	.149	.058	.190	.090	.087	.001
Substance Abuse & Adult Sexual Abuse	.074	.068	.186	-.018	.038	-.018	.135	.027	.064	.039
Substance Abuse & Adult Emotional Abuse (N=100)	.202*	.178	.058	.122	.092	-.035	.004	.055	.016	-.084

N=101, except where indicated; No. = number; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

Table 37 shows the correlations of each of these composites with the recidivism. Here, only two composites are related to outcome. The combination of substance abuse and adult sexual abuse shares a small significant negative correlation with the sum of aggregate probation sentences, suggesting that individuals with this abuse history tend to get shorter probation sentences. Of note, this could reflect less serious recidivism in that these women earn lighter probation sentences, or alternately that these women are earning other dispositions that are more serious (i.e., incarceration or conditional sentences) and thus actually reflect more serious outcomes. Likely, in this data set there are individuals in both circumstances given the approximately seven year follow up period, making this finding difficult to interpret.

The second composite that is related to recidivism, not surprisingly, is the combination of substance abuse and childhood emotional abuse. Specifically, the composite shares a small significant relationship with the presence of any, nonviolent, and violent reconvictions. Similarly, when we look at the inverse transform of the number of reconvictions another small significant relationship is revealed ($r=-.196$, $p=.050$).

Table 37: Correlations between the Gender Informed Composites and Recidivism.

Gender Informed Composites	Recidivism Present	No. of Reconvictions	Nonviolent Reconviction Present	No. of Nonviolent Reconvictions	Violent Reconvictions Present	Max of Time Sentenced	Sum of Aggregate Incarceration Sentences	Sum of Aggregate Probation Sentences
Single Parent	-.063	-.114	-.127	-.128	-.075	-.106	-.150	-.075
Substance Abuse & Child Physical Abuse	.078	-.010	.110	-.029	.180	.096	.040	-.016
Substance Abuse & Child Sexual Abuse	.033	.103	.029	.100	.070	.089	.093	.098
Substance Abuse & Child Emotional Abuse	.248*	.140	.247*	.127	.226*	.110	.097	.139
Substance Abuse & Adult Physical Abuse(N=100)	.159	.143	.164	.137	.176	.090	.066	.057
Substance Abuse & Adult Sexual Abuse	-.064	-.105	-.045	-.104	-.094	.027	-.056	-.230*
Substance Abuse & Adult Emotional Abuse(N=100)	.080	.184	.080	.186	.153	.102	.160	-.038

N=101, except where indicated; No. = number; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

Finally, when examining offence severity with the gender informed composites with the ordinal offence severity data (see Table 38), again, only the combination of substance abuse and childhood emotional abuse shares a small significant correlation.

Table 38: Spearman Correlations between the Gender Informed Composites and Offence Severity.

Gender Informed Variables	Min of Severity rank after assessment date	Max of Severity CCJS after assessment date
Single Parent	.062	-.094
Substance Abuse & Child Physical Abuse	-.081	.126
Substance Abuse & Child Sexual Abuse	-.041	.061
Substance Abuse & Child Emotional Abuse	-.202*	.217*
Substance Abuse & Adult Physical Abuse(N=100)	-.154	.176
Substance Abuse & Adult Sexual Abuse	.099	-.127
Substance Abuse & Adult Emotional Abuse(N=100)	-.066	.072

N=101, except where indicated; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

3.1.5.3. The incremental predictive validity of the gender informed variables and composites.

Next a series of stepwise logistic regression analyses was used to determine whether the gender-informed variables or their composites (analyzed initially in separate analyses to avoid multicolinearity concerns and then combined when appropriate into a final model) predict outcome over and above the risk instruments. The variables that were found to be significant in the earlier models testing the gender informed variables and their composites separately, but not found to be statistically significant in the final combined model are listed below each table depicting the results of the logistic regression analyses. Only the analyses that contained a gender informed variable or composite that significantly contributed to the predictive utility of the risk assessment instruments for an outcome variable are reported.

Table 39 depicts the model of the LSI-R predicting revocation. This model correctly classified 76.2 % of cases. Illegal financial support improves the predictive utility of the model (based on the log likelihood), but the individual variable only trends toward being a significant predictor of revocation according to the Wald statistic below. Further, and more importantly, the

confidence interval for illegal financial support crosses one suggesting that it is not a stable predictor as the direction of its relationship is not consistent. The composite of substance abuse and child physical abuse was selected in this model improving the prediction of the LSI-R for revocation over child emotional abuse, which was identified as significant in the gender informed variable only model.

Table 39: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Revocation with the LSI-R.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-6.644	1.520	19.097	1	.000	.001		
LSI-R	.178	.039	20.912	1	.000	1.195	1.107	1.290
Illegal Financial	2.060	1.137	3.281	1	.070	7.843	.844	72.845
Substance Abuse & Child Physical Abuse	-1.353	.648	4.360	1	.037	.258	.073	.920

Note: R^2 = .360 (Cox & Snell), .480 (Nagelkerke). Model $X^2(3) = 45.017$, $p < .001$. Excluded variable from previous analyses: Child Emotional Abuse, Wald (1) = 4.425, p =.035, *Exp (β)* = .240.

Table 40 depicts the model of the LSI-R predicting institutional misconduct. This model correctly classified 76 % of cases overall and 61% of the incidents of misconduct specifically. For institutional misconduct, child emotional abuse and adult emotional abuse variables, although significantly predictive variables when explored with the other gender informed variables alone, no longer improved the predictive utility of the LSI-R in this model which included the composite of substance abuse and child physical abuse and the composite of substance abuse and adult emotional abuse. However, as the confidence interval of the composite of substance abuse and child physical abuse crosses one, it is not a stable predictor as the direction of its relationship is not consistent. The presence of both substance abuse and a history of emotional abuse as an adult appears to decrease the risk of a women committing institutional misconduct (as reflected in Exp (β) values less than 1) and improves the prediction of the LSI-R with regard to this outcome.

Table 40: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Institutional Misconduct with the LSI-R.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β)	
							Lower	Upper
Constant	-4.459	.977	20.821	1	.000	.012		
LSI-R	.177	.039	20.884	1	.000	1.194	1.107	1.288
Substance Abuse & Child Physical Abuse	-1.232	.632	3.807	1	.051	.292	.085	1.006
Substance Abuse & Adult Emotional Abuse	-1.186	.594	3.990	1	.046	.305	.095	.978

Note: R^2 = .292 (Cox & Snell), .393 (Nagelkerke). Model $X^2(3) = 34.466$, $p < .001$. Excluded variables from previous analyses: Child Physical Abuse, Wald (1) = 3.807, p =.035, *Exp (β)* = .292; Adult Emotional Abuse, Wald (1) = 3.990, p =.046, *Exp (β)* = .305.

Table 41 depicts the model of the LSI-R predicting serious institutional misconduct. This model correctly classifies 81% of cases, but only 33.3% of serious misconduct incidents. Both childcare responsibility and the composite of substance abuse and child physical abuse improve the predictive utility of the LSI-R in predicting serious misconduct and their presence reflects a potential reduction in risk of reoffence with Exp (β) values less than 1. Because the substance abuse variable, found to be a significant predictor in the model examining gender informed variables only with the LSI-R, could not be tested in this model in addition to the composite variable of interest (due to the independence of errors assumption) it is not clear if the composite variable of substance abuse and child physical abuse (which was selected over the child physical abuse alone variable in this model) sufficiently captures the full potential contribution of substance abuse in the model of serious misconduct using the LSI-R.

Table 41: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Serious Institutional Misconduct with the LSI-R.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-3.660	1.080	11.476	1	.001	.026		
LSI-R	.128	.037	11.858	1	.001	1.137	1.057	1.223
Childcare Responsibility	-1.315	.582	5.110	1	.024	.268	.086	.840
Substance Abuse & Child Physical Abuse	-2.220	.740	8.997	1	.003	.109	.025	.463

Note: R^2 = .203 (Cox & Snell), .315 (Nagelkerke). Model $X^2(3) = 22.631$, $p < .001$. Excluded variables from previous analyses: Child Physical Abuse, Wald (1) = 5.171, p = .023, *Exp (β)* = .215; Adult Emotional Abuse, Wald (1) = 4.131, p = .042, *Exp (β)* = .284; Presence of substance abuse (not retested due to independence assumption), Wald (1) = 4.184, p = .041, *Exp (β)* = 14.334.

Table 42 depicts the model of the LSI-R predicting general recidivism. This model correctly classifies 70.3 of cases. As with the previous model of the LSI-R predicting revocations, illegal financial support is again an unstable predictor and the direction of its relationship to recidivism is unclear. Child sexual abuse appears to decrease one`s risk of recidivism and adds to the predictive utility of the LSI-R for general recidivism. Of note, child sexual abuse was selected over the composite of this form of abuse and substance use, suggesting the relationship between this variable and recidivism may be independent of substance abuse or that the contribution of substance abuse in predicting recidivism is sufficiently captured in the LSI-R score already contained in this model.

Table 42: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting General Recidivism with the LSI-R.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-3.879	1.092	12.610	1	.000	.021		
LSI-R	.111	.027	16.414	1	.000	1.117	1.059	1.178
Illegal Financial Support	1.533	.892	2.955	1	.086	4.630	.807	26.582
Child Sexual Abuse	-1.187	.533	4.965	1	.026	.305	.107	.867

Note: R^2 = .240 (Cox & Snell), .320 (Nagelkerke). Model $X^2(3) = 27.701$, $p < .001$. Excluded variables from previous analyses: Adult Sexual Abuse, Wald (1) = 2.838, p = .092, *Exp (β)* = .362; Substance Abuse and Child Sexual Abuse Composite, Wald (1) = 3.801, p = .051, *Exp (β)* = .329; Substance Abuse and Adult Sexual Abuse Composite, Wald (1) = 2.951, p = .086, *Exp (β)* = .358.

None of the gender informed variables or their composites improved the predictive utility of the LSI-R for violent recidivism. The LSI-R only models correctly classified 88.5% to 89% of cases. The R^2 values were between .099 and .100 (Cox & Snell) and .195 and .201 (Nagelkerke).

Table 43 depicts the model of the LS/CMI predicting revocation. This model correctly classifies 79.2 of cases. Having a spouse increases a woman's risk of revocation while the presence of both of substance abuse and child physical abuse in combination decrease her risk of revocation. Both of these variables improve the predictive accuracy of the LS/CMI on revocation. Interestingly, having a poor relationship with one's spouse is already captured on the LS/CMI as contributing to increased risk, as is being dissatisfied with being single. Of note, the composite substance abuse and child physical abuse was also a significant negative predictor in the model of the LSI-R and revocation.

Table 43: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Revocation with the LS/CMI.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-5.437	1.102	24.355	1	.000	.004		
LS/CMI	.230	.047	23.981	1	.000	1.258	1.148	1.379
Spouse	1.092	.534	4.179	1	.041	2.981	1.046	8.496
Substance Abuse & Child Physical Abuse	-1.313	.637	4.244	1	.039	.269	.077	.938

Note: R^2 = .385 (Cox & Snell), .541 (Nagelkerke). Model $X^2(3) = 49.169$, $p < .001$. Excluded variable from previous analyses: Child Emotional Abuse, Wald (1) = 2.706, p = .100, *Exp (β)* = .359.

Table 44 depicts the model of the LS/CMI predicting institutional misconduct. This model correctly classifies 80% of cases. In this model the composites of substance abuse and child physical abuse and adult emotional abuse both decrease a woman's risk of institutional misconduct (note, both of these composites were also included in the same model of institutional misconduct for the LSI-R, although the former variable was an unstable predictor). Further, while the composite of substance abuse and adult physical abuse improved the predictive utility of this model overall, its relationship with institutional misconduct is unstable and is it unclear whether it increases or decreases a woman's risk, given that its confidence interval crosses one. Of note, the composite variables included in this model were selected over the one previously significant gender informed variable, child physical abuse, suggesting that the combined contribution with substance use is especially important. Recall, substance abuse in general is also already captured in this model by the LS/CMI.

Table 44: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Institutional Misconduct with the LS/CMI.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β)	
							Lower	Upper
Constant	-4.149	.912	20.715	1	.000	.016		
LS/CMI	.193	.044	19.096	1	.000	1.213	1.113	1.323
Substance Abuse & Child Physical Abuse	-1.416	.686	4.260	1	.039	.243	.063	.931
Substance Abuse & Adult Physical Abuse	1.736	1.063	2.664	1	.103	5.673	.706	45.610
Substance Abuse & Adult Emotional Abuse	-2.143	1.029	4.338	1	.037	.117	.016	.881

Note: R^2 = .315 (Cox & Snell), .425 (Nagelkerke). Model $X^2(4) = 37.820$, $p < .001$. Excluded variable from previous analyses: Child Physical Abuse, Wald (1) = 3.667, p = .056, *Exp (β)* = .332.

Table 45 depicts the model of the LS/CMI predicting serious institutional misconduct. This model correctly classifies 80% of cases. Both childcare responsibility and the composite of substance abuse and child physical abuse decrease a woman's risk of serious institutional misconduct, and improve the overall prediction of the LS/CMI for serious institutional misconduct. Of note, while the composite of substance abuse and child physical abuse was selected for this model over child physical abuse alone, the presence of substance abuse in general was not tested in this model despite a previous significant relationship (given concerns about assumption violation). Thus, while the LS/CMI includes some measure of substance abuse, this model may be improved by further including another measure of substance abuse in general. This does not negate the importance of measuring child physical abuse in addition to substance abuse as both variables were selected in the gender informed variable only model as well. Note, this model of serious institutional misconduct is the same as that found for the LSI-R and serious institutional misconduct.

Table 45: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Serious Institutional Misconduct with the LS/CMI.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-2.823	.892	10.022	1	.002	.059		
LS/CMI	.119	.037	10.229	1	.001	1.127	1.047	1.212
Childcare Responsibility	-1.216	.562	4.681	1	.030	.296	.099	.892
Substance Abuse & Child Physical Abuse	-1.835	.695	6.974	1	.008	.160	.041	.623

Note: R^2 = .174 (Cox & Snell), .271 (Nagelkerke). Model $X^2(3)$ =19.146, $p < .001$. Excluded variable from previous analyses: Child Physical Abuse, Wald (1) = 4.875, p = .027, *Exp (β)* = .236; Adult Emotional Abuse, Wald (1) = 3.986, p = .046, *Exp (β)* = .291; Presence of Substance Abuse (not retested due to independence assumption), Wald (1) = 5.590, p = .018, *Exp (β)* = 18.413.

Table 46 depicts the model of the LS/CMI predicting general recidivism. This model correctly classifies 74.3 % of cases. The composite of substance abuse and child sexual abuse appears to decrease a woman's risk of recidivism, while the composite of substance abuse and child emotional abuse appears to increase a woman's risk of recidivism. Both significantly improve the predictive accuracy of the LS/CMI on general recidivism, as does the composite of substance abuse and adult sexual abuse. However, the composite of substance abuse and adult sexual abuse is not a stable predictor as evidenced by the confidence interval of its Exp (β) value crossing one, making the direction of its relationship with general recidivism unclear. This model differs significantly from the model of general recidivism with the LSI-R, which included no composite variables. Of note, while both of these risk assessment instruments contain very similar substance abuse domains, the operational definitions vary slightly and for this sample the LSI-R appeared to allow for the consideration of more historical substance abuse information when making ratings than did the refined LS/CMI rating instructions.

Table 46: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting General Recidivism with the LS/CMI.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-3.096	.756	16.779	1	.000	.045		
LS/CMI	.160	.038	17.716	1	.000	1.173	1.089	1.264
Substance Abuse & Child Sexual Abuse	-2.115	.813	6.768	1	.009	.121	.025	.594
Substance Abuse & Child Emotional Abuse	1.591	.777	4.198	1	.040	4.911	1.071	22.509
Substance Abuse & Adult Sexual Abuse	-1.103	.639	2.978	1	.084	.332	.095	1.161

Note: R^2 = .303 (Cox & Snell), .405 (Nagelkerke). Model $X^2(4) = 36.513$, $p < .001$. Excluded variables from previous analyses: Spouse, Wald (1) = 3.822, p = .051, *Exp (β)* = 2.769; Child Sexual Abuse, Wald (1) = 4.695, p = .030, *Exp (β)* = .290; Adult Sexual Abuse, Wald (1) = 4.504, p = .034, *Exp (β)* = .264.

Table 47 depicts the model of the LS/CMI predicting violent recidivism. This model correctly classifies 88.4 % of cases. Of note, none of the gender informed composites improved the predictive utility of the LS/CMI on violent recidivism. Childhood emotional abuse alone improved the model of the LS/CMI predicting violent recidivism; however, it is not a stable predictor as evidenced by the confidence interval of its Exp (β) value crossing one, making the direction of its relationship with violent recidivism unclear. Recall none of the gender informed variables or composites improved the predictive accuracy of the LSI-R on violent recidivism either.

Table 47: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables in Predicting Violent Recidivism with the LS/CMI.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-6.012	1.607	13.988	1	.000	.002		
LS/CMI	.104	.048	4.647	1	.031	1.110	1.010	1.220
Child Emotional Abuse	1.630	1.105	2.174	1	.140	5.101	.585	44.501

Note: R^2 = .123 (Cox & Snell), .242 (Nagelkerke). Model $X^2(2) = 12.622$, $p = .002$.

Table 48 depicts the model of the VRS predicting revocation. This model correctly classifies 71.9% of cases. None of the composites significantly improved the predictive utility of the VRS on revocation. Both the presence of substance abuse and illegal financial support increased a woman's risk for revocation and both variables improved the predictive utility of the VRS in this model of revocation. While the VRS contains an item addressing substance abuse, risk is only rated as increased if substance abuse contributed specifically to violence. The incremental value of the presence of substance abuse gender informed variable may reflect a more general relationship between substance abuse and revocation than is captured by the VRS. Further, as noted above the measure of illegal financial support may reflect a history of previous nonviolent criminal behaviour, which also would not necessarily be captured by the existing VRS items given its focus on the prediction of violence specifically.

Table 48: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables in Predicting Revocation with the VRS.

Included Variables							95% CI for Exp (β)	
	B	S.E.	Wald	df	Sig	Exp (β)	Lower	Upper
Constant	-4.667	1.301	12.866	1	.000	.009		
VRS	.042	.020	4.614	1	.032	1.043	1.004	1.084
Presence of Substance Abuse	1.679	.730	5.282	1	.022	5.358	1.280	22.424
Illegal Financial Support	2.270	1.118	4.121	1	.042	9.678	1.082	86.604

Note: R^2 = .268 (Cox & Snell), .359 (Nagelkerke). Model $X^2(3) = 29.996$, $p < .001$.

Table 49 depicts the model of the VRS predicting institutional misconduct. This model correctly classifies 76.2 % of cases. The composite of substance abuse and child physical abuse did not improve the prediction of the VRS on institutional misconduct as well as the addition of the child physical abuse variable alone. The presence of child physical abuse decreases one's risk of institutional misconduct as predicted by the VRS.

Table 49: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Institutional Misconduct with the VRS.

Included Variables							95% CI for Exp (β)	
	B	S.E.	Wald	df	Sig	Exp (β)	Lower	Upper
Constant	-2.548	.566	20.266	1	.000	.078		
VRS	.106	.022	22.595	1	.000	1.111	1.064	1.161
Child Physical Abuse	-1.171	.558	4.406	1	.036	.310	.104	.925

Note: R^2 = .284 (Cox & Snell), .383 (Nagelkerke). Model $X^2(2) = 33.793$, $p < .001$. Excluded variable from previous analyses: Composite of Substance abuse and Child Physical Abuse, Wald (1) = 4.467, p =.035, *Exp (β)* = .286.

Table 50 depicts the model of the VRS predicting serious institutional misconduct. This model correctly classifies 79% of cases overall, but only 28.6% of the cases with serious misconduct incidents. Note, the combined model did not include the presence of substance abuse variable to avoid an assumption violation. When the combined model without the presence of substance abuse was examined, childcare responsibility, a significant predictor in the previous model was also no longer selected for inclusion in this model as statistically relevant. As the only composite variable included in this combined model (composite of substance abuse and child physical abuse) did not significantly add to the predictive utility of the VRS on serious institutional misconduct, the earlier model testing the full list of gender informed variables likely yields the best model for predicting serious institutional misconduct with the VRS. See Table 51 for this model.

Table 50: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Serious Institutional Misconduct with the VRS.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-2.424	.647	14.044	1	.000	.089		
VRS	.099	.025	15.228	1	.000	1.104	1.050	1.160
Child Physical Abuse	-2.006	.727	7.611	1	.006	.134	.032	.559
Adult Physical Abuse	-1.390	.679	4.188	1	.041	.249	.066	.943

Note: R^2 = .218 (Cox & Snell), .339 (Nagelkerke). Model $X^2(3)$ =24.558, $p < .001$. Excluded variable from previous analyses: Composite of Substance Abuse and Child Physical Abuse, Wald (1) = 9.283, p = .002, *Exp (β)* = .096; Childcare Responsibility, Wald (1) = 4.185, p= .041, *Exp (β)* = .261; Presence of Substance Abuse (not retested due to independence assumption), Wald (1) = 4.940, p = .026, *Exp (β)* = 16.574.

Table 51 depicts the model of the VRS predicting serious institutional misconduct. This model correctly classifies 84.4 % of cases overall, but only correctly classifies 47.6% of the cases with serious misconduct incidents. However, this is an improvement over the combined model in Table 50 without the presence of substance abuse and childcare responsibility. While the presence of substance abuse reflects an increase in risk of serious misconduct over that predicted by the VRS, childcare responsibility, child physical abuse and adult physical abuse all appear to reflect a decrease in risk of serious institutional misconduct.

Table 51: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables in Predicting Serious Institutional Misconduct with the VRS.

Included Variables							95% CI for Exp (β)	
	B	S.E.	Wald	df	Sig	Exp (β)	Lower	Upper
Constant	-3.130	1.123	7.763	1	.005	.044		
VRS	.078	.028	7.859	1	.005	1.081	1.024	1.141
Presence of Substance Abuse	2.808	1.263	4.940	1	.026	16.574	1.393	197.134
Childcare Responsibility	-1.342	.656	4.185	1	.041	.261	.072	.945
Child Physical Abuse	-1.727	.732	5.564	1	.018	.178	.042	.747
Adult Physical Abuse	-1.831	.787	5.416	1	.020	.160	.034	.749

Note: R^2 = .290 (Cox & Snell), .446 (Nagelkerke). Model $X^2(5)$ =32.901, $p < .001$.

Table 52 depicts the model of the VRS predicting general recidivism. This model correctly classifies 74.3 % of cases overall. Illegal financial support increased a woman's risk of recidivism over and above that predicted by her VRS score, while a history of child sexual abuse decreases her risk. Again, illegal financial support captures a history of nonviolent criminal behaviour that may not be included in the exisiting VRS items. Further, as with previous models of the VRS and institutional misconduct the gender informed abuse variables were selected in the final model as better contributors than the composites.

Table 52: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting General Recidivism with the VRS.

Included Variables							95% CI for Exp (β)	
	B	S.E.	Wald	df	Sig	Exp (β)	Lower	Upper
Constant	-3.951	1.104	12.806	1	.000	.019		
VRS	.117	.026	20.639	1	.000	1.124	1.069	1.183
Illegal Financial Support	2.199	.953	5.318	1	.021	9.013	1.391	58.395
Child Sexual Abuse	-1.746	.621	7.893	1	.005	.174	.052	.590

Note: R^2 = .326 (Cox & Snell), .435 (Nagelkerke). Model $X^2(3)$ =39.922, $p < .001$. Excluded variable from previous analyses: Composite of Substance Abuse and Child Sexual Abuse, Wald (1) = 6.659, p = .010, *Exp (β)* = .122; Composite of Substance Abuse and Child Emotional Abuse, Wald (1) = 3.909, p= .048, *Exp (β)* = 4.533; Composite of Substance Abuse and Adult Sexual Abuse, Wald (1) = 3.404, p = .065, *Exp (β)* = .285.

None of the gender informed variables or their composites increased the predictive utility of the VRS on violent recidivism. The models with the VRS only correctly classified 89.6% to 90% of cases overall, but only 27.3% of cases with violent recidivism. Recall there were only 11 cases of violent recidivism in the entire sample.

Table 53 depicts the model of the SIR-R1 predicting revocation. This model correctly classifies 72% of cases overall. Note, as none of the gender informed abuse variables were predictive in earlier models, the presence of substance abuse variable did not need to be excluded from the combined model to avoid a violation of the independence assumption. Interestingly the composites of substance abuse and child and adult physical abuse did add to the predictive utility of the SIRR1 on revocations better than the presence of substance abuse. The composite of substance abuse and child physical abuse decreased a woman's risk of revocation as predicted by the SIR-R1 as seen in previous models with other risk instruments and outcomes, while the composite of substance abuse and adult physical abuse reflected an over fivefold increase in risk of revocation. Illegal financial support also contributed to increased risk of revocation over and above that predicted by the SIR-R1 alone. Finally, recall that the lower a score on the SIR-R1 the higher one's risk of recidivism, hence the expected *Exp (β)* value less than one reflecting a negative relationship between risk score and risk of revocation.

Table 53: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Revocation with the SIR-R1.

Included Variables							95% CI for Exp (β)	
	B	S.E.	Wald	df	Sig	Exp (β)	Lower	Upper
Constant	-2.051	1.231	2.776	1	.096	.129		
SIRR1	-.170	.044	15.312	1	.000	.843	.774	.918
Illegal Financial Support	2.556	1.151	4.933	1	.026	12.882	1.350	122.874
Substance Abuse & Child Physical Abuse	-1.352	.659	4.207	1	.040	.259	.071	.942
Substance Abuse & Adult Physical Abuse	1.688	.618	7.452	1	.006	5.409	1.610	18.176

Note: R^2 = .352 (Cox & Snell), .470 (Nagelkerke). Model $X^2(4)$ =43.420, $p < .001$. Excluded variable from previous analyses: Presence of Substance Abuse, Wald (1) =4.195, p = .041, *Exp (β)* = 4.600.

Table 54 depicts the model of the SIR-R1 predicting institutional misconduct. This model correctly classifies 74% of cases overall, but only 53.7% of cases where institutional misconduct was present. Further, while both the composites of substance abuse and adult physical and adult emotional abuse contribute to improving the predictive utility of the model of institutional misconduct as predicted by the SIR-R1, neither composite is stable. That is, the confidence intervals of their *Exp (β)* values cross one and thus the direction of their relationship with institutional misconduct is unclear.

Table 54: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Institutional Misconduct with the SIR-R1.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β)	
							Lower	Upper
Constant	-.021	.425	.003	1	.960	.979		
SIRR1	-.095	.031	9.358	1	.002	.909	.855	.966
Substance Abuse & Adult Physical Abuse	1.549	.847	3.344	1	.067	4.707	.895	24.761
Substance Abuse & Adult Emotional Abuse	-1.488	.824	3.258	1	.071	.226	.045	1.136

Note: R^2 = .178 (Cox & Snell), .240 (Nagelkerke). Model $X^2(3) = 19.641$, $p < .001$. Excluded variables from previous analyses: Adult Emotional Abuse), Wald (1) = 2.927, p = .087, *Exp (β)* = .401; Presence of Substance Abuse (not retested due to independence assumption), Wald (1) = 5.001, p = .025, *Exp (β)* = 5.184.

Table 55 depicts the model of the SIR-R1 predicting serious institutional misconduct. This model correctly classifies 80% of cases overall, but only 19% of cases where serious institutional misconduct was present. Note, the combined model did not include the presence of substance abuse variable to avoid singularity. However, as the only composite variable included in this combined model did not significantly add to the predictive utility of the SIR-R1 on serious institutional misconduct, the earlier model testing the full list of gender informed variables likely yields the best model for predicting serious institutional misconduct with the SIR-R1. (This was also true for the VRS model of serious institutional misconduct.) See Table 56 for the model of the SIR-R1 and gender informed variables predicting serious institutional misconduct.

Table 55: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Serious Institutional Misconduct with the SIR-R1.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	.076	.475	.025	1	.873	1.079		
SIRR1	-.072	.033	4.696	1	.030	.931	.872	.993
Childcare Responsibility	-1.014	.528	3.686	1	.055	.363	.129	1.021
Child Physical Abuse	-1.229	.584	4.435	1	.035	.293	.093	.918

Note: R^2 = .113 (Cox & Snell), .176 (Nagelkerke). Model $X^2(3) = 12.018$, $p = .007$. Excluded variables from previous analyses: Composite of Substance Abuse and Child Physical Abuse, Wald (1) =4.775, p = .029, *Exp (β)* = .259; Adult Emotional Abuse, Wald (1) =3.933, p = .047, *Exp (β)* = .291; Presence of Substance Abuse (not retested due to independence assumption), Wald (1) =7.870, p = .005, *Exp (β)* = 28.847.

Table 56 depicts the model of the SIR-R1 predicting serious institutional misconduct using only the gender informed variables. This model correctly classifies 84.4% of cases overall, but only 42.9% of cases where serious institutional misconduct was present. However, this is an improvement over the combined model above. Substance abuse appears to be the strongest predictor of serious institutional misconduct of the gender informed variables and its presence reflects a more than 28 time increase in likelihood of serious institutional misconduct as predicted by the SIR-R1. In this model, child physical abuse and adult emotional abuse each decrease a woman's risk of serious institutional misconduct. However, while childcare responsibility appears to improve the predictive utility of the model of the SIR-R1 predicting serious institutional misconduct, this variable is unstable as the confidence intervals of its *Exp (β)* value cross one and thus the direction of their relationship with serious institutional misconduct is unclear. Moreover, and most importantly, the confidence intervals of its *Exp (β)* value for the SIR-R1 also crosses one, suggesting in combination with these variables it is not a stable predictor of serious institutional misconduct. This is not entirely surprising given that a clear steplike progression from one risk level to the next was not seen for general or violent recidivism survival analyses.

Table 56: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables in Predicting Serious Institutional Misconduct with the SIR-R1.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	-2.127	1.139	3.488	1	.062	.119		
SIRR1	-.015	.041	.137	1	.711	.985	.909	1.067
Presence of Substance Abuse	3.362	1.198	7.870	1	.005	28.847	2.754	302.145
Childcare Responsibility	-1.180	.603	3.824	1	.051	.307	.094	1.003
Child Physical Abuse	-1.267	.637	3.958	1	.047	.282	.081	.981
Adult Emotional Abuse	-1.234	.622	3.933	1	.047	.291	.086	.986

Note: R^2 = .230 (Cox & Snell), .354 (Nagelkerke). Model $X^2(5) = 25.131$, $p < .001$.

Table 57 depicts the model of the SIR-R1 predicting general recidivism using only the gender informed variables. This model correctly classifies 81.2% of cases overall. Again, while illegal financial support, child sexual abuse, and the composites of substance abuse and child physical and emotional abuse improve the predictive utility of this model, the confidence interval of their *Exp (β)* values cross one and thus the direction of these relationships with general recidivism is unclear.

Table 57: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting General Recidivism with the SIR-R1.

Included Variables	B	S.E.	Wald	df	Sig	Exp (β)	95% CI for Exp (β) Lower	Upper
Constant	.700	.973	.518	1	.472	2.015		
SIRR1	-.234	.054	19.174	1	.000	.791	.712	.879
Illegal Financial Support	1.463	.931	2.473	1	.116	4.321	.697	26.772
Child Sexual Abuse	-1.380	.786	3.084	1	.079	.252	.054	1.174
Substance Abuse & Child Physical Abuse	-1.518	.907	2.800	1	.094	.219	.037	1.297
Substance Abuse & Child Emotional Abuse	1.419	.892	2.529	1	.112	4.132	.719	23.737

Note: R^2 = .382 (Cox & Snell), .509 (Nagelkerke). Model $X^2(5) = 48.536$, $p < .001$. Excluded variables from previous analyses: Spouse, Wald (1) =2.750, p = .097, *Exp (β)* = 2.531; Adult Sexual Abuse, Wald (1) =2.637, p = .104, *Exp (β)* = .318.

Finally, Table 58 depicts the model of the SIR-R1 predicting violent recidivism. This model correctly classifies 88.1% of cases overall, but 0% of the cases with violent recidivism present. Again, while childhood emotional abuse appears to improve the overall utility of the model, the confidence interval of its *Exp (β)* value crosses one and thus the direction of this relationship with violent recidivism is unclear.

Table 58: Logistic Regression Analyses: Relative Contributions of Gender Informed Variables and Composites in Predicting Violent Recidivism with the SIR-R1.

Included Variables							95% CI for Exp (β)	
	B	S.E.	Wald	df	Sig	Exp (β)	Lower	Upper
Constant	-3.330	1.029	10.467	1	.001	.036		
SIRR1	-.099	.046	4.615	1	.032	.906	.827	.991
Child Emotional Abuse	1.887	1.091	2.995	1	.084	6.603	.779	55.988

Note: R^2 = .107 (Cox & Snell), .214 (Nagelkerke). Model $X^2(2) = 11.401$, $p = .003$. Excluded variable from previous analyses: Composite of Substance Abuse and Child Emotional Abuse, Wald (1) =2.308, p = .129, *Exp (β)* = 3.559.

4.1 Discussion

This study examined the predictive utility of four traditional risk assessment instruments (i.e., LSI-R, LS/CMI, VRS, SIR-R1) as well as the predictive utility of a number of gender informed variables for a variety of recidivism outcomes. The results are promising with regard to the utility of the traditional risk assessment instruments with a sample of federal female offenders. They also underscore the importance of a number of gender informed variables for this population, as well as the complexity of the relationships between these variables and their combined contribution to predicting various forms of recidivism. While this study was unable to clearly define which gender informed variables may be most useful as supplements to the traditional risk assessment tools, it did highlight that each of the traditional risk assessment instruments was improved by at least one gender informed variable when predicting a variety of recidivism outcomes, with the exception of violent recidivism. This discussion will address basic descriptive statistics for these risk assessment instruments, and psychometric properties, predictive utility and incremental predictive utility of these risk instruments and gender informed variables. In additional to contextualizing these findings in the existing literature, implications for correctional practice, limitations of this study, and future directions are discussed.

4.1.1 Basic descriptive statistics.

The utility of the risk assessment instruments included in this study for the federal female offender population received little or no previous attention. Thus, understanding the comparability of base rates of reoffending for this sample and the similarity in the scores obtained is critical for understanding and building upon the existing risk prediction data, especially if one plans to make risk statements based on our existing understanding of what a particular score means.

4.1.1.1 Recidivism rates.

In the current sample, the violent recidivism rate (11%) was comparable to a previous study of federal female offenders (Folsom & Atkinson, 2007), although the rates of general recidivism (52%) were higher for the current sample than previously reported. This is likely attributable to the longer follow up time in the current study. Most other studies of female offenders to date have focused on provincial samples which for male offenders have been shown to have higher rates of recidivism than their federal counterparts particularly in the Medium, High, and Very High LS/CMI risk groups (Andrews, Bonta, & Wormith, 2004). This base rate

difference appears to apply to federal and provincial female offenders as well as a previous study with provincial samples showed higher recidivism rates than currently observed despite a shorter follow up time, especially for institutional (as compared to community sanctioned) offenders (Rettinger, 1998). Similarly, although the current study does not compare male and female offenders, it is important to also consider gender differences in base rates of reoffence when examining the utility of purportedly gender neutral risk instruments. Andrews and colleagues (2011b) found that the mean base rate of recidivism was substantially lower for females (.29) than males (.50) across five data sets and noted that, as is often warned by feminist researchers, over-prediction may be occurring with female offenders in the lower risk groups of the LS/CMI. Finally, Hardyman and Van Voorhis (2004) also noted that women's rates of serious institutional misconduct were four times lower than their male counterparts with the same custody rating, and can lead to reduced freedoms and increased confinement. Thus, while differences in follow up time and setting make it challenging to directly compare base rates, it does appear that federal female offenders may have lower recidivism rates than incarcerated provincial offenders and than their male federal offender counterparts, which highlights that risk statements about this population need to reflect these differences.

4.1.1.2 Mean risk assessment scores.

Secondly, beyond examining the base rates of recidivism, examining the mean risk scores for the women in this sample also yielded some important differences from the previous literature which are likely attributable to methodological differences, as well as population heterogeneity based on setting (institution versus community). Specifically, the mean LSI-R score in the current study was 28.29 (*SD* = 10.92), while the only other study of federal female offenders to date (Folsom & Atkinson, 2007) using the LSI-R produced a mean score of 18.0 (*SD* = 10.5), despite warning that their sample had an overrepresentation of high security level women. This difference is likely due to the computer administered versus clinician rating method for completing the LSI-R, despite research supporting the validity of self-report measures for predicting recidivism and institutional misconduct (see Loza, Neo, Shahinfar, & Loza-Fanous, 2005, for a comparison of the LSI-R to the Self-Appraisal Questionnaire for female offenders and Walters, 2006, for a meta-analysis with both genders). Provincial female offenders and American felony female offenders also had a lower LSI-R score (*M* = 18.74, *SD* = 9.21; Rettinger & Andrews, 2010; M = 17.18 and 18.56, Holtfreter, Reisig, & Morash, 2004) than

observed in the current sample, but this difference was smaller when examining provincial institutional offenders alone (*M* = 24.28, *SD* = 8.03; Rettinger, 1998) or would-be incarcerated offenders (*M* = 25.05, SD = 7.24; Lowenkamp, Holsinger, & Latessa, 2001). Of note, the mean LSI-R score for the women in the latter sample was also not significantly different from a sample of American felony male offenders in the same study. Although, one study of all women admitted to the state department of corrections over a three month period by Salisbury, Van Voorhis, and Spiropoulos (2009) reported a higher LSI-R total score (33.4: *SD* = 7.6) than was observed in the current study, it would appear that federal female offenders score similarly to other incarcerated women (and possibly men) on the clinician rated LSI-R, but not female offenders in the community or whose ratings rely on self report. This is consistent with the observed differences in base rates of recidivism described above, and provides more support for interpreting these scores in similar way.

In exploring the mean LS/CMI score (*M* =22.51, *SD*=9.143), it is clear that this sample of female offenders had comparable scores overall to those typically found for male and female offenders in provincial institutional setting (*M*=22.39, *SD*=8.47 and *M* =21.79, *SD*=8.94, respectively, Andrews, Bonta, & Wormith, 2004), which provides some support for not creating separate risk level cutoff scores for each gender. Moreover, the scores for the current sample of federal female offenders were more comparable to those of other institutional offenders than those previously reported for federal male offenders, whose LS/CMI scores had a mean of 25.50 (recalculated from the raw LSI-R data, *SD* = 8.17; Andrews, Bonta, & Wormith, 2004), which was three points higher on average than their federal female counterparts in the current study, and separate norms were not suggested for these men. In contrast, Andrews and colleagues (2011b) have recently suggested that explorations of different cutoff scores for males and females may be warranted for the LS/CMI. In their most recent examination of multiple mixed gender data sets, they found that women tended to be overrepresented in the very low and low risk categories and underrepresented in the high risk categories (they did not include a separate very high risk category in their study). They also confirmed that females were recidivating substantially less than males in the moderate and lower risk levels, but not in the high risk cases. However, many of the offenders included in these samples were young offenders and women in the community not incarcerated adult women, which the current study would suggest may look more like other institutional offenders.

As previously noted, the current study is one of the first to examine the VRS on a sample of adult female offenders. While, risk scores did not significantly differ for male and female youths on the VRS-YV (Stockdale, 2008), the youth version of the VRS, the current study found mean VRS scores (M = 25.25, *SD*= 14.91) much lower than previously reported for other adult male offenders. The VRS authors (Wong & Gordon, 2006) reported mean VRS score of 41.86 (*SD* = 16.42) in a large male sample also comprised mainly of federal offenders with varied criminal histories. A second study by the authors of the VRS with a sample of male violence-prone offenders who had been selected for a high intensity treatment program yielded an even higher mean score (*M* = 55.23, *SD* = 10.70; Wong, Gordon, & Gu, 2007). Other research reflected mean scores ranging from 35.01 (*SD* = 10.11; Grevatt, Thomas-Peter, & Hughes, 2004) in a small sample of male psychiatric inpatients in England, most of whom had violent convictions to 60 (*SD* = 9.1; Sheldon & Krishnan, 2009) for patients residing on a Dangerous and Severe Personality Disorder unit. While setting and population differences appear to contribute to varied mean scores, using a file review only method did not appear to affect scores. Despite past literature suggesting file reviews could yield more conservative ratings (Wong, 1988), a sample of medium security mentally disorder male patients in the United Kingdom produced a mean VRS score of 41.0 (*SD* = 11.3; Dolan & Fullam, 2007) using file only ratings also which is very comparable to the VRS authors' first study results.

A second study that appeared to expand this sample of 136 male patients by including 11 women patients from the same setting found the same mean VRS score, but suggested there was a trend for the female offenders to have slightly lower mean scores on the static, but not dynamic components of the VRS (Dolan, Fullam, Logan & Davies, 2008). However, these conclusions are based on a very small number of women who may differ significantly from other female correctional populations especially given the variability seen in the male samples highlighted above, making it premature to rule out more significant gender differences as suggested by the current data. Thus, more research will be required to understand gender differences that may exist with regard to VRS scores in the larger population of female offenders.

Finally, the mean SIR-R1 score was 5.12 (*SD* = 8.173) for this sample. Neither Nafekh and Motiuk (2002) or Bonta, Pang, and Wallace-Capretta (1995) provided a mean SIR score for their federal female offender samples; however, the range of scores for the latter study was

comparable (-18 to +15 in their study compared to -17 to +19 in the current study). Mills and Kroner (2006) examined the SIR scale in male federal offenders and reported a mean reversed score of 3.9 (S.D. = 8.8, thus a higher score represents increased risk) which would be much higher (i.e., more than one standard deviation) than the mean in the current study if it were reverse scored. While this suggests a gender difference is likely with regard to the magnitude of scores, more research is again recommended to explore a possible difference directly if this instrument is to be pursued for use with female offenders.

Overall, while setting and methodological differences make it challenging to compare risk scores across samples, the current data lend support to the large empirical literature suggesting the LSI-R and LS/CMI appear to yield similar mean scores to other populations of incarcerated male and female offenders. While past gender differences with VRS scores for girls and women were largely not found, at the least the latter of these examinations was on a very small sample of women, and the data from the current study suggest that such differences are likely at least for federal female offenders. Moreover, while SIR-R1 scores in the present study are similar to other federal female offender samples, they likely differ from other male offenders. This is not surprising as the SIR-R1 relies heavily on static factors and the static factors on the VRS were the ones that tended to reflect a gender difference even with only a few cases examined (Dolan, Fullam, Logan & Davies, 2008). Thus, additional attention is warranted with regard to potential gender differences for the VRS and SIR-R1 to ensure that risk statements based on our existing understanding of what a particular score means are modified to reflect the lower mean scores earned by women.

4.1.2 Psychometric analyses.

Interrater reliability and internal consistency of the risk assessment instruments as well as the gender informed variables were explored to ensure that these constructs could be measured reliably and to examine their shared content. The interrelatedness of the risk assessment instruments and gender informed variables were also examined to begin to look at the degree to which unique information was being captured by the gender informed variables. Overall, these analyses also highlighted some methodological limitations and related practical considerations related to the quality of the data available.

More specifically, as predicted, the risk instruments had good interrater reliability overall; however, some individual items did not. This seemed largely attributable to the quality

and consistency of information available in the women's electronic files. These issues point to a number of possible concerns with regard to the information being collected and considered with evaluating female offenders. Specifically, while information about many criminogenic risk/needs areas are routinely collected, data about some of the potentially relevant factors contributing to success while incarcerated and upon release (e.g., history of illegal financial support; housing/accommodation) may not be. Other researchers have noted similar limitations (Rettinger & Andrews, 2010) Thus, it may be helpful to provide some guidance about these relevant factors to front line staff to help guide their assessment procedures and understanding of these women. Routinely including a risk assessment instrument in the assessment process may be a way to ensure other relevant risk and need variables are explored.

Further, when there are differing experiences with or opinions of the women's behaviours and attitudes (e.g., a quiet woman in a group setting is disengaged or contemplative; a woman understands the factors that contribute to her offence cycle or she is making excuses to justify her offending), at times little attention was paid in the file information to integrating and explaining these differences. Perhaps with the opportunity to interview these women when amalgamating all of the available information into a comprehensive risk assessment, this integration would be possible. It seems that a more integrated case conceptualization of each woman would be of benefit to all treatment providers and front line staff when interacting with these women on a regular basis. To this end, staff may routinely discuss their impressions of the women with whom they work without this information being added to the electronic files. However, integration of this information into the women's files may help with continuity of care, particularly as they make the transition from the institution to the community. This information may also be especially important for understanding a women's mental health functioning and motivation/readiness for treatment. Reconciling differences in file information is critical for creating a reliable data source about these women.

Internal consistency was also strong for the risk assessment instruments, with the exception of the SIR-R1. This scale had a Cronbach's alpha of .699, which was closer to Nafekh and Motiuk's (2002) Cronbach's alpha value of .77 for the measure's targeted population of male non-Aboriginal federal offenders than the values achieved by the other risk instruments in this study. Past research on these instruments also reflected equally high internal consistency, although like with the SIR-R1, static components were found to have lower alpha values

(Andrews & Bonta, 1995; Andrews, Bonta, & Wormith, 2004; Wong & Gordon, 2006). Moreover, the gender informed variables were not expected to have strong internal consistency as their inclusion in their study was to reflect a diverse group of relevant factors for female offenders that would not necessarily reflect a shared construct. The complexity and diversity of the relationships between these gender informed variables was also reflected in subsequent analyses related to recidivism to be discussed further below.

When comparing the four risk assessment instruments in this study to one another they were highly correlated with the strongest relationship existing between the LSI-R and the LS/CMI as expected given the overlap in their item content ($r = .880$ in the current study and $r = .96$ to $.98$ in the LS/CMI manual, Andrews, Bonta, & Wormith, 2004). The weakest relationship was between the SIR-R and VRS ($r = -.575$) which is also not surprising give the lack of dynamic items in the former instrument and the focus on violence (as opposed to general offending) in the latter. The strong relationship between the LSI-R and VRS ($r = .734$) provides evidence for the utility of the VRS in predicting recidivism with female offenders, given the support for the utility of the LSI-R with other female offenders in general (Lowenkamp, Holsinger, & Latessa, 2001; Folsom & Atkinson, 2007; Smith, Cullen, & Latessa, 2009; Rettinger & Andrews, 2010).

Finally, when comparing the gender informed variables and risk assessment instruments, the latter had largely moderate correlations with all forms of childhood abuse and adult physical abuse (except with the SIR-R1), but not with adult sexual abuse, which may be attributable to the small reported incidence of this abuse in this sample. Recall each of these instruments, except the SIR-R1, measures the quality of marital relationships was well as substance abuse which have been suggested in combination with abuse history to be predictive of future reoffending and thus some of their shared variance already may be captured in the risk scores (Lowenkamp, Holsinger, & Latessa, 2001). Further, adult emotional abuse had a small correlation with both the LSI-R and the VRS, but not the other two risk instruments, which is interesting as the LSI-R and VRS are the only two risk assessment instruments that directly assess mental health concerns, which may be contributing to this relationship. This is especially relevant as the development of mental health or emotional concerns is one of the proposed ways which abuse history may lead to offending behavior (Lowenkamp, Holsinger, & Latessa, 2001; Salisbury & Van Voorhis, 2009) although this pathway has previously focused on childhood

abuse, while relationship difficulties were seen as contributing to increased risk of abuse as an adult and thus more recidivism (Salisbury & Van Voorhis, 2009).

Overall, these psychometric analyses underscored methodological and practical concerns related to the quality and content of file information and the importance of a reliable and consistent data source. They also highlighted the interrelatedness of the items of the risk assessment instruments, as well as the diversity of the gender informed variables. The risk assessment instruments overlap largely with one another, while some relationships are also seen with some of the gender informed variables, especially those related to abuse history. While abuse history items are not included in the traditional risk assessment tools, some variables like substance abuse, marital relationship and mental health concerns which combine with abuse history in a number of pathways (Daly, 1992; Lowenkamp, Holsinger, & Latessa, 2001; Salisbury & Van Voorhis, 2009) are included and may capture the effect of these trajectories to some degree as part of the total risk/need score.

4.1.3. Predictive validity.

The current study set out to begin to explore the utility of the VRS in predicting recidivism for female offenders, to expand the existing literature on the LSI-R and LS/CMI to a federal sample of female offenders and to clarify the utility of the SIR-R1 with this population given past inconsistency. Further, while each tool was expected to be a valid predictor of a variety of outcomes, the aim was also to determine if any risk assessment instrument predicted any outcomes better than the other risk assessment instruments. It was hypothesized that the LS/CMI and LSI-R would best predict institutional misconduct, revocations, and predict general recidivism. Moreover, the VRS was expected to have the strongest relationships with violent recidivism and offence severity.

As expected, all of the risks assessment instruments were significantly related to general and violent recidivism, institutional misconduct and revocation. The predictive validity of the instruments was replicated across a number of outcome measures (i.e., prevalence and frequency of various types of offence/incident, offence/incident severity, and sentence length measures), and these relationships reflected moderate to large effects in general consistent with past literature.

Specifically, a large meta-analysis of a variety of female offender samples found an average correlation of .35 between LSI-R scores and general recidivism which may be somewhat

smaller than in the current sample of federal female offenders (r = .42), especially given the large sample size captured in the meta-analysis (N = 14,737, with 27 effect sizes, Smith, Cullen, & Latessa, 2009). Unlike a past study of a self report version of the LSI-R in federal female offenders, in the current study a significant relationship with violent offending was found and the observed relationship with general recidivism was again larger (r = .42 as compared to r = .30, Folsom & Atkinson, 2007). The AUC values for the current sample were also somewhat larger than previously found for federal female offenders (.74 as compared to .67 for general recidivism, and .78 as compared to .67 for violent recidivism, Folsom & Atkinson, 2007).

Further, past research on the LS/CMI, yielded higher correlations with general and violent recidivism (r = .63 as compared to r = .47, and r = .45 as compared to r = .30, respectively) and higher AUC values (.87 as compared to .77 and .86 as compared to .77, respectively; Rettinger & Andrews, 2010) for provincial offenders than in the current study. However, the effects remain large in general despite the smaller current sample. The values were slightly more comparable to a large scale study combining a variety of female offenders samples which yielded a correlation with general recidivism of .53 and an AUC of .827 (Andrews, et al., 2011b).

With regard to the VRS, the correlations found in the current study with a variety of outcome measures were comparable with past research on the VRS for general and violent recidivism in male offenders (Wong & Gordon, 2006). They were somewhat higher with regard to institutional misconduct than in a mixed gender sample (ranging from .32 to .51 in the current study as compared to .28 and .37 in Dolan, Fullam, Logan, & Davies, 2006), which may be due in part to different operational definitions of misconduct as they focused exclusively on institutional aggression/violence. Moreover, AUC values observed in the present study, which ranged from .76 (recent serious misconduct) to .84 (violent recidivism) on the variety of outcome measures, were higher than those reported with regard to institutional aggression, which ranged from .63 to .69 in Dolan, Fullam, Logan, and Davies (2006) and was .71 in Dolan and Fullam (2007), and with regard to recidivism in male offenders, which ranged from .71 to .75 in Wong and Gordon (2006).

In the present study the AUC values for general (.83) and violent recidivism (.77) were again somewhat higher than in past research on the SIR-R1 (.77 and .72, respectively, Nafekh & Motiuk, 2002). The correlation with general recidivism (-.54) was much higher than in past

research (.25; Bonta, Pang, & Wallance-Capretta, 1995). These samples also examined federal female offenders, but reported some systematically missing data (number of dependents and employment items), which may partially account for the lower validity in the earlier studies as these factors may be especially relevant to female offenders.

The risk assessment instruments also reflected significant negative relationships with time to first non-violent reconviction (except the SIR-R1, which shared a significant positive relationship, as expected) suggesting that higher risk individuals recidivated more quickly than their lower risk peers. However, time to first violent reconviction was not related to any of the risk instrument scores. This may reflect that higher risk individuals do not recidivate any more quickly than lower risk individuals, but this null result may also be due to the small incidence of violent recidivism in this study, and thus replication is critical before interpreting this finding, especially given the paucity of comparable research on this outcome measure.

When looking at the comparative predictive validity of the tools, the LS/CMI and LSI-R were expected to have the strongest relationships with institutional misconduct. The LS/CMI was the best predictor of the presence of institutional misconduct during the index offence, except for the VRS. Similarly, the VRS was the best predictor of the frequency of institutional misconduct during the index offence compared to all the other risk assessment instruments except the LS/CMI. In both circumstances the LSI-R did not significantly differ from the second best predictor, suggesting that the small significant differences observed between these three instruments, likely did not reflect a clinically relevant difference. This may be true for the SIR-R1 scale as well, as although the LS/CMI and VRS routinely outperformed the SIR-R1 scale (for serious misconduct incidents as well) the SIR-R1 scale was not outperformed by the LSI-R. Previous meta-analytic research has found static risk assessment instruments, like the SIR-R1, to be better predictors of institutional violence than dynamic risk assessment instruments like the other risk assessment instruments in the current study (Campbell, French, & Gendreau, 2009), although even serious misconduct incidents in this study may not constitute violence. Of note, institutional misconduct during the index offence reflects more concurrent validity than predictive validity, as the file information for the assessment ratings was in part reflecting the same time period as the occurrence of some of the institutional misconduct. Further, subsequent analyses of institutional misconduct following the assessment period only yielded no differences between the risk assessment instruments, suggesting that this more proximal measure of

antisocial behavior also was not better predicted by one tool over another. This is not entirely surprising as past research comparing multiple risk assessment instruments on institutional misconduct also did not find any significant differences between the predictive utility of the risk assessment instruments (Kroner & Mills, 2001).

The LS/CMI and LSI-R were expected to have the strongest relationships with revocations, as well. Although, both of these instruments outperformed the VRS on the presence of revocations and the LS/CMI outperformed the VRS on the frequency of revocations, the other instruments were not differentiated from each other in their predictive utility. Thus, these statistically significant differences do not clearly identify a superior risk assessment instrument when predicting revocations (based on these comparisons alone). Again this is consistent with past research on a variety of risk assessment instruments (Kroner & Mills, 2001).

Further, while the LS/CMI and the LSI-R were expected to best predict general recidivism, no reliable differences between the predictive utility of any of the tools was observed for the presence or frequency of general recidivism. While the lack of statistical difference is consistent with previous research comparing a number of risk assessment instruments (Kroner & Mills, 2001), including meta-analytic research (Gendreau, Little, & Goggin, 1996; Olver, Stockdale, & Wormith, 2009), unlike past research on male offenders (Gendreau, Little, & Goggin, 1996), the LSI-R (and related LS/CMI) had slightly lower magnitude correlations than the other risk assessment instruments (VRS and SIR-R1) in this study when examining the presence of reconvictions. Surprisingly, the SIR-R1 has the strongest correlation with both the presence and frequency of reconvictions, despite having been frequently a poor predictor for misconduct and revocation.

Finally, while the VRS was expected to have the strongest relationships with violent recidivism and offence severity, there were no significant differences between the risk assessment instruments when examining the strength of their predictive utility. However, as noted above, the low incidence of violent reconvictions in this sample makes it difficult to draw more definitive conclusions about the comparative utility of these risk assessment instruments in predicting violent recidivism based on correlational data in women despite evidence for the general utility of each risk assessment instrument for this purpose. (Recall, the AUC values, which are unaffected by low baserates, for the risk assessment instruments and violent recidivism were all high.) Further, past research also found little differentiation between various risk

assessment instruments when predicting violent recidivism (Campbell, French, & Gendreau, 2009; Olver, Stockdale, & Wormith, 2009; Yang, Wong, & Coid, 2010), especially, when making within study comparisons (Gendreau, Goggin, & Smith, 2002). However, past meta-analytic research has highlighted both higher and lower effect sizes for a variety of risk assessment instruments predicting violent recidivism for female offenders relative to their male counterparts, suggesting that gender differences in the predictive utility of these risk assessment instruments may exist (Yang, Wong, & Coid, 2010).

While the risk assessment instruments may not vary statistically in their ability to predict various outcome measures of recidivism in this sample of federal female offenders, there are other relevant functions of the risk assessment instrument which may guide instrument selection (Yang, Wong, & Coid, 2010). Namely, some other benefits of the risk assessment instruments that may inform instrument choice include the inclusion of dynamic variables that may guide treatment (Gendreau, Little, & Goggin, 1996; Kroner & Mills, 2001; Campbell, French, & Gendreau, 2009), the inclusion of a measurement of change/treatment readiness to guide therapeutic approach (Wong & Gordon, 2006; Wong, Gordon, & Gu, 2007), or consideration of the context and clinical questions being answered by the assessment (Gendreau, Goggin, & Smith, 2002).

Overall, the strength of the predictive utility of each of the risk assessment instruments appeared comparable and at times stronger than past research for other male and female offenders despite the small sample size examined. Small significant differences observed between the risk assessment instruments likely did not reflect clinically relevant differences in their predictive utility. This is promising for the utility of each of these tools for women offenders in general and provides stronger support for their use with federal female offenders more specifically. This study also underscored the consistency of the predictive utility of the risk assessment measures across a variety of outcome definitions (e.g., prevalence, frequency, severity, offence type) in general. Finally, as the risk assessment instruments do not differ on their predictive utility, the other relevant functions of the tools would best guide risk assessment instrument selection for federal female offenders.

4.1.4 Predictive validity of risk levels.

In addition to examining the predictive utility of the risk instrument scores, looking at the utility of the risk levels of each instrument in differentiating offenders' risk for recidivism is

important. Whether recidivism rates increase as risk level increases similarly to other populations suggests whether unique cut off scores or norms may be warranted more so than comparison of the mean scores of individual groups alone. Further, as discussed above, understanding the recidivism rates at each risk level is critical when including risk statements in clinical assessments to ensure that risk prediction is not misunderstood. However, it is important to note that with a small sample size such as that of the current study, examinations that further divided the sample into smaller groups made having sufficient power to find these differences difficult. This was especially true for the violent recidivism comparisons given the low incidence of violent recidivism in the current sample, which was half that of their federal male counterparts (Bonta, Rugge, & Dauvergne, 2003). Thus a steplike progression from one risk level to the next was sought whereby there was an increase in the rate of recidivism as the risk levels increased.

4.1.4.1 General recidivism.

Women in the high risk groups on each of the traditional needs/risk assessment instruments were predicted to have higher and faster rates of general recidivism than those with lower scores, as illustrated by survival analysis. While there was differentiation with the LSI-R for the two lowest groups relative to the three highest groups on time to general recidivism in the expected direction, the five risk level differentiation suggested in the manual was not produced. While Manchak, Skeem, Douglas, and Siranosian (2009) also did not find that the risk levels suggested in LSI-R manual predicted time to recidivate (generally) in a sample of women who had been convicted of serious violent offences, they were able to find a steplike progression with the data trichotomized. Further, Folsom and Atkinson (2007) compared a three level differentiation which would be approximately equivalent to the low and low/moderate risk groups and the remaining three risk levels combined. As with the current study, they also found a two level differentiation, but with the low risk group differing significantly from the other two. Recall that the mean risk score in the Folsom and Atkinson's study was much lower than in the current sample which may explain a lower level differentiation and in fact be quite comparable to the current findings. In a larger sample, Rettinger (1998) also suggested that the five LSI-R risk levels were distinguishable in her sample of provincial offenders both in the community and in an institution; however, it was not clear if each level was distinct despite an overall significant difference and an observed steplike increase in recidivism relative to risk level. Thus, while a steplike progression from low to higher risk groups appears to be generally present for LSI-R

scores with female offenders, more data would be required to determine if the five level differentiation is possible and if cut off scores need to be modified.

Better differentiation was observed for the LS/CMI where no women in the low risk group reoffended, and risk scores also clearly distinguished the moderate risk group from the high and very high risk groups. Further, the high and very high risk groups overlapped only to a small degree relative to the large confident intervals observed, despite the small frequencies in these groups. While, the proposed five level differentiation was not produced, this was due in part to no women scoring in the very low risk group which may relate to the women in the current sample being more serious offenders than the provincially sentenced women who were included in the normative sample. Even in the normative sample only 2.2% of provincial inmates scored in the very low risk level (Andrews, Bonta, & Wormith, 2004). Rettinger (1998) also found some support for a significant relationship and significant difference between risk levels and recidivism for the LSI-OR (almost identical to the LS/CMI) although, as mentioned above, these analyses did not seem to ensure distinct differentiation at every level. The fact that the normative sample for the LS/CMI included a much larger group of female offenders may have contributed to a three (or potentially four) level differentiation observed in the current study as opposed to the two levels found with its predecessor the LSI-R for this sample of federal female offenders. This level of differentiation adds additional support to the utility of the LS/CMI for federal female offenders to classify federal female offenders into meaningful risk level groups.

Moreover, when comparing the recidivism rates of the federal female offenders in the current study to a diverse group of male and female offenders (including provincial, community, and young offenders; Andrews et al., 2011b) the recidivism rates of the high/very high risk groups appear comparable which is good support for the current cutoff scores for these groups. However, the low and moderate risk groups had lower rates of recidivism in the current sample than both the male and female offender samples (Andrews et al., 2011b). Lower frequencies were also observed in the current sample for these risk groups, which is consistent with differences in recidivism base rates observed between provincial and federal offenders. Although gender differences were not explored directly in the current study (i.e., there was no male comparison group), Andrews and colleagues (2011b) have already identified the possible need to modify the cutoff scores in the lower and moderate risk levels as outlined above to ensure that recidivism rates are comparable. The current findings would support this change to ensuring that

risk statements are not misinterpreted when describing lower and moderate offenders from different populations, but no such change appears to be necessary for the higher risk offenders from different populations who are already reliably classified by the current LS/CMI risk level cut offs.

With regard to the VRS, three unique risk levels were evident for this sample of federal female offenders when examining general recidivism. This is consistent with past research with male offenders (Wong & Gordon, 2006) and to some degree a mixed sample of young offenders (Stockdale, 2008) where there was clear differentiation for all levels except only a trend between the high and medium risk level groups. While a three level differentiation suggests the VRS has similar utility to the LS/CMI when predicting general recidivism, it is important to recall that unlike the LSI-R and LS/CMI, the risk levels for the VRS were created based on truncating the current data, not predetermined cut off scores, resulting in a statistical advantage. Further, targeting fewer risk levels to differentiate is also more easily done with a small sample although concerns about the power of these comparisons remain. This notwithstanding, the VRS does appear to be quite promising in its ability to differentiate federal female offenders based on their risk levels when predicting general recidivism and meaningful risk descriptors could be produced for this population assuming these results are replicable. Of note, no studies to date have compared male and female offenders risk of recidivism based on the VRS, thus the need for potential gender specific risk level cut off scores remains to be seen, but would be suggested by the considerably lower mean risk score observed in the current study.

Finally, the SIR-R did not differentiate between risk levels appropriately as a steplike progression from highest to lowest risk was not produced. This is consistent with limitations of this risk assessment instrument observed in some past research on the utility of this instrument with female offenders (Bonta et al., 1995) who also failed to observe a steplike progression in a small sample (N = 81), although Nafekh and Motiuk (2002) reported the SIR-R1 accurately distinguishes between the five risk groups in their larger sample of federal female offenders (N = 342). This lack of reliable differentiation, despite a strong correlation with general recidivism, implies difficulties with the prescribed cut off scores and significantly limits the utility of this measure, especially as the lowest risk level is the only one which can be reliably differentiated as currently defined. That being said, given the ease at which this tool is completed and its ability to identify low risk offenders in this sample more reliably, it may have utility as for screening

instrument for determining who may be identified as warranting a more through risk appraisal or triage to the community. Given the alternative purpose proposed for the SIR-R1, the comparability of the recidivism rates at each risk level to male offenders is less relevant.

While the small sample size of the current study limited the power of risk level comparisons for the risk assessment instruments, the data do tend to reflect a steplike progression for each of the instruments, except the SIR-R1 when predicting general recidivism. The current data provide good support for the use of the LS/CMI risk levels for federal female offenders, although research to date suggests that some modification of the cut-off scores may be required to make recidivism rates more comparable at lower and moderate risk level for male and female offenders. Thus, clinical risk statements should acknowledge such potential differences. The LSI-R data were not as strong for the exiting cut off scores, and the LS/CMI risk levels appear to have benefited from a larger and more diverse normative sample of female offenders. While the VRS also showed good risk level differentiation, this is the first study to use this instrument to predict recidivism in female offenders and thus predetermined risk level cut off scores still need to be created using a larger data set of female offenders. Finally, given the ease use of the SIR-R1 scale, but its repeatedly demonstrated limitations in producing a steplike progression from low to higher risk levels, its utility for female offenders appears limited to acting as a screening instrument as low risk female offenders are the only ones reliably identified with the current cut off scores.

4.1.4.2 Violent recidivism.

With regard to violent recidivism, as noted above, those women with higher risk scores on the risk assessment instruments were predicted to have higher and faster rates of violent recidivism than those with lower scores, as illustrated by survival analysis. For the LSI-R a three level differentiation was observed. However, unlike with the two level differentiation observed for general recidivism, there was better differentiation within the three highest groups. Specifically, the high risk group was distinct, while the moderate and medium/high groups' confidence intervals remained overlapping. None of the women who scored in the low or low/moderate risk levels recidivated violently, thus the two lowest groups remain undifferentiated as they did for general recidivism. Rettinger (1998) also found no violent recidivists in her lowest two risk levels on the LSI-R in her sample of incarcerated provincial offenders, and found a positive correlation between the five LSI-R risk levels and violent

recidivism, but did not appear to explore whether each level was distinguishable from all the others. The LSI-R risk level groups were not differentiated in Folsom and Atkinson's (2007) study when predicting violent recidivism which is not surprising as they did not find a significant relationship between risk score and violent recidivism in general either. The trend in the current data for only higher risk offenders to receive new violent reconvictions in steplike progression from lower to higher risk supports the utility of the LSI-R risk levels for federal female offenders.

In the current study there was a two level differentiation found for violent recidivism with the LS/CMI risk levels, as all recidivists fell in the high or very high risk levels which were not distinguishable from each other, and no one in low or medium risk groups was reconvicted for a violent offence. Recall, no woman was rated as scoring in the very low risk group, thus conclusions cannot be drawn about this risk level. While this suggests that the lower and midrange scores are not under-classifying violent recidivists, it is important to recall that there were nonrecidivists in the high and very high risk level groups and thus prediction was not perfect (nor would it be expected to be). Again, Rettinger (1998) found a significant relationship between violent recidivism and LSI-OR risk levels with no violent recidivists in the lowest two risk levels, but did not examine whether each risk level was distinct from the others. Thus, as with the LSI-R, there is a trend for violent recidivists to be classified in the higher risk levels, as would be expected further supporting the utility of the risk levels of the LS/CMI in general also.

With regard to the VRS, two unique risk levels were evident for this sample of federal female offenders when examining violent recidivism. Specifically, none of the women in the lowest risk level group had a violent reconviction, but the high and medium risk level groups' confidence intervals overlapped slightly. While this is the only risk assessment instrument specifically designed to predict violence, this two level differentiation is promising given the low incidence of violent recidivism in the sample. Past research with male offenders (Wong & Gordon, 2006) and a mixed sample of young offenders (Stockdale, 2008) both demonstrated clear differentiation for three risk levels. As previously mentioned past research examining risk levels has created these cut-off scores based on the data being examined and it may be beneficial to move toward creating cut-off scores that could be replicated in other samples, as the interpretation of a risk level is often sought to be applied at the level of an individual being assessed making a reference sample critical. To this end, it is promising the violent recidivists

are being classified in the higher risk groups as would be expected. Given that the mean scores observed in the current study of federal female offenders are much lower than those found for men (see discussion above), these cut-offs will likely need to be unique for each gender.

Finally, the risk level groupings for the SIR-R1 again perform most poorly relative to the other risk assessment instruments for predicting violent recidivism. Although there are three differentiated levels, the low risk group contains recidivists while the low to moderate risk level does not and the three highest risk levels are not distinguishable from each other. Nafekh and Motiuk (2002) reported the SIR-R1 accurately distinguishes between the five risk groups in their sample of federal female offenders for violent recidivism using ROCs, although it is not clear that this difference was examined at each risk level specifically and may reflect an overall difference only. In the current sample, the lack of steplike progression is more evident for violent than general recidivism. Thus, the SIR-R1 does not even appear to be promising as a screen to identify the need for further assessment for violent recidivism, as even at the lowest risk level, the progression toward higher recidivism rates is not present.

Overall, while few of the risk assessment instruments differentiated all of the intended risk levels when examining violent recidivism, the LSI-R, LS/CMI and VRS appeared promising as they showed an appropriate progression from lower risk with no recidivism to higher risk with more recidivism. Further, the limited differentiation is also likely due in part to the small sample size (and especially the related small cell size at some risk levels) more than the utility of the instruments given the moderate to strong correlation and AUC values produced by the continuous data. Also, there was some risk level overlap only at extremes of large confidence intervals, which is likely due to in part the long and variable follow up time of the women in this sample and the metric used to reflect this time period (days). This was especially true for the VRS. Further, while cutoff scores may need to be adjusted to ensure more similar general recidivism rates between men and women at lower risk levels, the expected progression of risk level and increased violent recidivism further supports past findings that high risk offenders of both genders are being similarly classified in general (Andrews et al., 2011b).

4.1.5. Incremental predictive validity.

Several gender informed variables and composites of these variables were examined to determine if and to what degree they improved the predictive validity of the four risk assessment instruments. While there has been some support in the literature for these gender informed

variables, which included childcare responsibility (especially being a single parent), abuse/victimization history (alone and in combination with substance abuse), financial challenges, and self harm experiences, to be potential criminogenic needs which could improve the prediction of recidivism in various forms, the findings have not been consistent (Bonta, Pang, & Wallace-Capretta, 1995; McClellan, Farabee, & Crouch, 1997; Rettinger, 1998; Lowenkamp, Holsinger, & Latessa, 2001; Wright, Salisbury, & VanVoorhis, 2007; Salisbury & Van Voorhis, 2009; Salisbury, Van Voorhis, & Spiropoulos, 2009).

In the present study, the presence and intensity of substance abuse did reliably correlate with most outcomes (but not violent recidivism); however, it is a criminogenic need already identified by the dynamic risk assessment instruments being examined. In this study, its potential relevance as a gender informed variable was to form composite variables in combination with the abuse history variables to examine their incremental validity. Further, substance abuse had a small to moderate correlation with a variety of abuse experiences in childhood and adulthood as expected. Of note, adulthood physical abuse was the most frequent form of abuse history reported in the sample, occurring for 68% of women. Further, self harm/suicidality was also related to both substance abuse and abuse history in general. These relationships provide some support for pathway models that have looked at these variables in combination (McClellan, Farabee, & Crouch, 1997; Wright, Salisbury, & VanVoorhis, 2007; Salisbury & Van Voorhis, 2009) in that these variables do appear to co-occur in female offenders.

However, while the gender informed variables were expected to have significant predictive validity for violent and general recidivism, revocation, institutional misconduct, and severity of reoffence, these relationships were few and of a small magnitude. History of adult sexual abuse showed some promise with regard to predicting the presence of institutional misconduct, and possibly shorter probation sentences (which is an ambiguous outcome that could reflect less severe probation sentencing or more frequent alternative sentences, likely custodial sentences). Childhood sexual abuse was also related to the presence of revocation. The frequency of minor misconduct incidents and presence of violent reconvictions were related to childhood emotional abuse. Salisbury, Van Voorhis, and Spiropoulos (2009) also found that childhood abuse was related to serious misconduct in prison, while a trend was present for the effect of adulthood emotional abuse. Adult abuse variables, but not childhood abuse variables,

also predicted rearrest in their study. Self harm/suicidality only positively related directly to violent reconvictions and possibly sentence length in the current study, while previous research has linked it to general recidivism (Bonta, et al., 1995). Further, without consistency between multiple definitions of the outcome constructs (e.g., presence, frequency, severity) it is difficult to ascertain if these few correlations (among so many) are more likely reflective of type 1 error. Moreover, of all the gender informed variables and outcomes only childhood emotional abuse and violent recidivism yielded a significant AUC. However, the lack of significant relationships between these variables and a variety of recidivism outcomes may not be inconsistent with the pathways literature in that the proposed utility of these variables is not when their effect is examined in isolation, but in combination.

Further, the variables included in this study may not have tapped the intended constructs. For example, although it did not yield a significant AUC, history of illegal financial support appeared to be more consistently related to a number operational definitions of recidivism and thus is more likely than the other gender informed variables to be a reliable predictor (not surprisingly as it is by definition a previous criminal act). History of social assistance and illegal financial support, both thought to be measuring economic pressures, were not correlated suggesting that they are measuring distinct constructs. Thus, perhaps the latter is more reflective of criminal behavior in general than a gendered economic motivation underlying financially profitable crimes. Interestingly, to further differentiate these constructs, history of illegal financial support was only related to the composite of substance abuse and child emotional abuse, while history of social assistance was related to all other abuse composites except adult sexual abuse (which had a small reported incidence). This differentiation may also support the pathways theory whereby female offenders who are economically motivated are more similar to their criminally oriented male counterparts, while women with histories of substance abuse, and victimization as children or adults are on other gendered pathways to crime (Daly, 1992; Reisig, Holtfreter, & Morash, 2006). However, given the low interrater reliability of a history of illegal financial support, these findings should be replicated when the quality of the data can be better assured.

The gender informed composites examining substance abuse and abuse history in combination showed more promise in the current study. The composite of substance abuse and childhood sexual abuse had small correlations with both measures of revocation, as did the

composite of substance abuse and adult physical abuse, which also showed promise for predicting any and minor misconduct incidents, especially when distributions were normalized. Similarly, the composite of substance abuse and childhood emotional abuse showed the most promise for predicting (albeit still to a small degree) any and minor misconduct incidents at various times, as well as the presence of revocations. Further, the composite of substance abuse and childhood emotional abuse was the only one to reliably predict recidivism, demonstrating small correlations with the presence of any, nonviolent and violent reconvictions and both offence severity scales.

This is consistent with past research that revealed that emotional maltreatment appeared to have stronger relationship to psychological functioning in the long term than other forms of abuse (Kaplan, Pelcovitz, & Labruna, 1999). Specifically, emotional abuse has been shown to be a stronger predictor than physical abuse of childhood physical aggression, delinquency, violence victimization and perpetration, suicide attempts, depression, anxiety, trauma symptoms, inpatient hospitalization, and internalizing and externalizing behaviours in general possibly due to its negative impact on self-esteem and affect regulation (Vissing, Straus, Gelles, & Harrop, 1991, Mullen, Martin, Anderson, Romans, & Rerbison, 1996; McGee, Wolfe, & Wilson, 1997; Spertus, Yehuda, Wong, Halligan, & Seremetis, 2003; Berzenski &Yates, 2010). Alternately, it has been suggested that all forms of abuse may produce a generalized or nonspecific disruption in development that may be difficult to distinguish in the long term (Mullen, et al., 1996) and it may be that emotional abuse, "one of the most destructive and pervasive forms of abuse," is a component of all other forms of abuse (p. 2, Wright, 2007). Thus, the robust effect of childhood emotional abuse relative to other forms of abuse found in past literature may explain why it was the form of abuse that was most consistently related to a variety of recidivism outcomes in the current study.

When examining the predictive utility of the gender informed variables and composites over and above the risk assessment instruments, the results varied greatly and few variables consistently improved the prediction across the various outcomes or risk assessment instruments. However, this variability may affirm the complexity of the contribution of these variables, and underscore that there is remaining variability to be explained in a variety of outcomes that is not captured by the current risk assessment instruments (as most models did benefit from the addition of at least one gender informed variable or composite). Further, as suggested above, the

interrelatedness of these variables may also account for the inconsistencies in the results if it is the shared variance between the variables or composites that is contributing to the predictive utility, especially for the abuse variables/composites.

More specifically, illegal financial support and the composite of substance abuse and childhood physical abuse were the gender informed variables/composites that most frequently improved the prediction of revocation across a number of the risk assessment instruments. Illegal financial support increased risk of revocation, while the composite of substance abuse and child physical abuse appeared to reduce women's risk of revocation. Interestingly, the LSI-R's predictive utility was improved by both of these gender informed constructs despite already measuring some domains thought to be related. For example, history of illegal financial support would be expected to overlap with both or either of the criminal history and financial domains of the LSI-R, while it was expected that the composite of substance abuse and abuse history would overlap to some degree with both the substance abuse and personal/emotional domains (neither of which are found to *reduce risk*). Again, this underscores the importance of examining these gender informed constructs in more detail to understand their unique contribution more fully. For example, it may be that women with such histories are seeking counseling or mental health treatment upon release to cope with the residual impact of these experiences and thus have more support upon release to avoid revocations (unfortunately, such information was not collected as part of this study.) However, replication will also be important as Salisbury, Van Voorhis, and Spiropoulos (2009) added a child abuse variable to the LSI-R in addition to variables reflecting mental health and relationships and did not find an improvement in the prediction of technical violations, but did argue it improved the prediction of serious misconduct.

With regard to misconduct and serious misconduct the composite of substance abuse and childhood physical abuse again showed promise in predicting a reduction in the risk of misconduct over that predicted by the LS/CMI and LSI-R to a lesser degree, as did the composite of substance abuse and adulthood emotional abuse. Moreover, a history of childhood physical abuse alone also predicted a decreased risk of misconduct for the VRS and to a lesser degree the SIR-R1, while substance abuse also added positively to the predictive utility of these tools for serious misconduct. This difference in the selection of both the gender informed variables over the composite may reflect that the LSI-R and LS/CMI already measure substance abuse in general, while the VRS and SIR-R1 do not. Interestingly, Wright, Salisbury, and VanVoorhis

(2007) did not find substance abuse related to institutional adjustment in their sample of female offenders, acknowledging this was not consistent with past research and may have reflected better institutional control of substance abuse-related misconduct. The relevance of child abuse to institutional adjustment is consistent with their findings as well as VanVoorhis, Salisbury, Wright, and Bauman (2008) who listed child abuse as a risk factor for institutional adjustment difficulties, but as a responsivity factor for their probation and parole samples.

Finally, childcare responsibility appeared to be a promotive factor over and above the remaining risk assessment instruments for serious misconduct (the logistic regression analyses highlighted that the SIR-R1 scale was not a reliable predictor of serious institutional misconduct), but not misconduct in general. In contrast, Bonta, Pang, and Wallace-Capretta (1995) found that single mothers were more likely to recidivate, and thus were at increased risk, although they did not examine institutional misconduct specifically, and this finding was not replicated in the current study either with regard to general recidivism. Further, Wright, Salisbury, and VanVoorhis (2007) found parental stress marginally related to institutional adjustment difficulties early during the incarceration, but not later in the sentence and they did not focus on child custody stressors which may account for some of this difference. Thus, it may be more important to explore the nature of parent child relationships in order to understand the impact of childcare responsibility more clearly both while in custody and upon release. These relationships may be both motivating factors for desistance or stressors that contribute to a return to criminal behavior, which, not surprisingly, merely classifying female offenders as having childcare responsibility failed to capture.

As for general recidivism, a number of gender informed variables and composites improved the predictive utility of the risk assessment instruments, but only a history of childhood sexual abuse reliably improved more than one model (that of the LSI-R and VRS) and contributed to a decreased risk of general recidivism. Further a history of illegal financial support also improved the prediction of recidivism (in a positive direction) for the VRS, which is not surprising as the risk assessment tool does not capture nonviolent criminal histories or financial concerns specifically. Moreover, the LS/CMI's predictive utility was increased positively by the composite of substance abuse and childhood emotional abuse, and negatively by the composite of substance abuse and child sexual abuse (a potential promotive factor). In contrast, Rettinger and Andrews (2010) did not find that childhood or adulthood abuse variables

had a significant relationship with recidivism once the women's LS/CMI score was examined. Further, as with the current study, even their initial small correlations between the abuse variables and recidivism were all in the positive direction and thus did not suggest any reduction in recidivism for women with such histories. However, Bonta, Pang, and Wallace-Capretta (1995) did also find that women who were abused physically as adults were less likely to recidivate providing some support for a related finding. That said, as the potential for promotive factors only arose from the combination of both multi-domain risk assessment tools and gender informed variables, it seems that a much better understanding of the interplay of these variables is required to understand this potential effect.

Finally, none of the gender informed variables reliably improved the predictive utility of the risk assessment instruments for violent recidivism. The lack of improvement on the current risk assessment instruments for violent recidivism may be at least in part reflective of the low incidence of violent recidivism in this sample. That said, childhood emotional abuse alone (although not a stable predictor) showed promise in that it added to the predictive power of models with the LS/CMI and SIR-R1 scales predicting violent reconviction. It also correlated with a number of outcomes on its own, and was the only significant AUC of the gender informed variables and thus should be explored further.

Overall, the current findings suggest that the risk assessment instruments can be improved by the gender informed variables or composites when predicting most forms of recidivism as the logistic regression models of the risk assessment instruments were consistently improved by at least one of the gender informed variables or composites. Further, while the gender informed variables alone were not predictive of the recidivism outcomes for the most part, the composites, especially those related to substance abuse and abuse history, were much more promising as would be expected based on the pathways theory. However, there was limited consistency in the result to explain which forms of abuse reliably contributed to improving the risk assessment instruments or each particular outcome. Childhood emotional abuse was the most promising in that it shared the most consistent relationships with outcome alone, in combination with substance abuse, and showed some promise for increasing the predictive utility of the risk assessment instruments for general and violent recidivism. This is consistent with past research that suggested that childhood emotional abuse is more predictive of longer term psychological outcomes than other forms of abuse (see Kaplan et al., 1999 for a review). As

well, given the interrelatedness of the gender informed variables and the supposition that emotional abuse is a potent component of all other forms of abuse (Wright, 2007) it may be the most appropriate target of future research. Finally, understanding how the these experiences in childhood develop into increased risk for recidivism will be critical for guiding improved risk prediction and targeting appropriate treatment for these women.

4.1.6. Limitations and future directions.

Several limitations existed in the current study. Specifically, the small sample size, especially when subgroup comparisons (e.g., risk level groupings) were examined, limits the statistical power of the data. However, the representativeness of the sample and the thorough sampling procedures do help support the generalizability of the findings in this study to a larger population of federally sentenced women. Further, the infrequent incidence of violent recidivism limits the conclusions about the utility of these measures when predicting violence, but does highlight areas for further exploration as the current findings were promising.

Replication of the utility of these risk assessment instruments for federal female offenders is critical as only a few studies to date have examined this population and some differences from other female samples have been observed (e.g., base rates of recidivism, different mean scores), despite much research to date on measures like the LSI-R and LS/CMI. There has been little research to date supporting the use of the SIR-R1 scale with female offenders, except Nafekh and Motiuk (2002) who may not have looked for a steplike progression in risk and recidivism at each unique level, which appears to be the risk assessment instrument's most significant limitation. Thus, future research may want to focus on its potential as a screening tool for nonviolent offenders, given the ease and speed at which it can be completed, or upon refining the items included in the scale. Further, this is only the second study to score the VRS for female offenders (given cursory examination of 11 female offenders by Dolan, Fullam, Logan, & Davies, 2008) and the first to examine its predictive validity for recidivism outcomes and thus the application of this risk assessment instrument specifically, although promising, remains in its infancy.

Additionally, all of the risk assessment instruments were completed by the same rater. Despite efforts to rate each measure independently, only one file review was completed, thus some bias in ratings is possible as the rater was primed to look at data relevant to the other risk assessment instruments as well. Counter-balancing the ratings would not necessarily control for this effect, as attention is paid to information on additional variables during the file review based

on the requirements of the other risk assessment instruments. Moreover, interrater reliability does provide some support that another rater could achieve similar results; however, the raters used the same process of completing one file review to score all of risk instrument and the gender informed variables. This method has some ecological validity, in that most clinicians completing multiple risk assessment instruments would also only likely review the file once, but it would be very unlikely to rate all four of these risk instruments on one client/patient. Thus, replicating the current findings with independent raters would be beneficial to ensure bias from one instrument to another was controlled, especially as comparisons between the instruments are made.

Also, although past research has suggested that file review only methods for risk assessment research are acceptable (Wong, 1988; Gendreau, Goggin, & Smith, 2002; Campbell, French, & Gendreau, 2009), relying solely on file review limited the scope and depth of the assessment information available in this study. This was especially true for the gender informed variables whose operational definitions were limited to the information likely recorded on file on a routine basis. This meant relying more heavily on self-report for some constructs as well, like history of victimization. Also, these variables were often limited to the presence of such concerns at any point in time, and thus were not explored in a way that reflected any possible dynamic nature of the underlying constructs. Risk assessment research has clearly demonstrated that looking at dynamic variables in addition to static variables improves predictive ability and allows treatment targets to be identified. Given the support in the larger risk assessment literature for the benefits of measuring constructs in ways that reflect their fluidity and their temporal proximity to the outcomes being predicted (Andrews & Bonta, 2003; Hardyman & Van Voorhis, 2004), refining our operational definitions beyond static or historical gender informed variables would likely eliminate some sources of error variance in future analyses and potentially outline more specific treatment targets for these women. This may be especially true given the complex and overlapping relationships between these gender informed variables as highlighted by their presence in multiple pathways in Daly's original model (e.g., childhood abuse and substance abuse are each in more than one of Daly's 1992 proposed pathways). For example, the proximal impact or influence of the women's abuse histories was not considered in the current study, just that such experiences were endured in the past. This does not reflect the degree to which the women had or had not healed from these experiences or any lasting consequences that were still present in their lives. More attention should be focused on the process of how these factors

continue to influence reoffence long after their occurrence. For example, some constructs like self-efficacy/self-esteem may be more proximal effects of abuse histories and have been linked to recidivism (Task Force on Federally Sentenced Women, 1990; VanVoorhis, Salisbury, Wright, and Bauman, 2008; Van Voorhis, Wright, Salisbury, & Bauman, 2010). These constructs were not explored in the present study as they were too difficult to measure reliably from file information alone. Thus, including an interview component may be especially helpful when identifying a group of dynamic gender informed variables for future research.

Further, the current study focused on only completing a one-time risk assessment instrument rating of only the portion of the risk assessment instruments that yielded a risk/need total score. This is not reflective of the recommended procedures for two of the risk assessment instruments in particular. Namely, one unique strength of the LS/CMI is that it guides the clinician through an additional case management and responsivity assessment, that although not scored may impact future decisions about placement, treatment, and other recommendations for offenders. The utility of these sections was not explored in the current study, and may be an important consideration when selecting one risk assessment instrument over another (Andrews, Bonta, & Wormith, 2004). Future research should examine the utility of these case management sections more thoroughly in this population.

Similary, the impact of VRS stages of change ratings were not examined as each file was only rated at one time point. As the ability to measure change is one of the aspects most unique to the VRS (Wong & Gordon, 1998 – 2003; Wong & Gordon, 2006; Wong, Gordon, & Gu, 2007), further research exploring the utility of change scores resulting from examining pre and post ratings will be especially interesting to pursue, given the promise demonstrated in the current study for use of the VRS with female offenders.

The current study only examined the utility of these tools in total and did not explore the relative utility of individual items or domains. Past research has suggested that some risk/need domains share stronger or weaker relationships with recidivism for women than for men and thus by definition are gender sensitive/salient not neutral (e.g., frequency of substance abuse, see McClellan et al., 1997; attitudes and associates, see Van Voorhis, Wright, Salisbury, & Bauman, 2010; associates, financial, family/marital and criminal history, see Heilbrun, et al., 2008; substance abuse Andrews, et al., 2011b). This would be interesting for future research to examine with regard to federal female offender samples for each risk assessment instrument, and

for female offenders in general with regard to the VRS given the paucity of data to date on this instrument comparing gender. This may provide guidance for reweighting or adding/deleting items in specific domains to improve predictive utility or confirm the current structure of these instruments.

In addition, although the current study compared its findings of federal female offenders to past literature on a variety of populations, having a specific federal male offender comparison group would have allowed for direct examination of possible gender differences. This is especially important as other methodological differences between studies limit the validity of such comparisons. Further, while much data exist on the risk assessment instruments, few studies have compared the predictive utility of the gender informed variables for men and women especially in federal offenders (see Brown & Motiuk, 2005 for a review of some related constructs) and none have compared the incremental predictive validity of these constructs to the risk assessment instruments as in the current study.

Moreover, while the utility of these risk assessment measures and gender informed variables have been explored in the prediction of recidivism, their utility in identifying appropriate treatment targets for federal female offenders should also be explored. Future studies should aim to determine if use of a risk assessment instrument to match interventions to a specific women's needs yields more effective correctional treatment based on the principles of risk, need, and responsivity. Further, as few of the gender informed variables shared direct relationships with recidivism outcomes in the current study, they should be examined as potential responsivity factors in future research that allows for the additive effect of addressing these variables over and above treating criminogenic needs to be examined.

This type of examination of the traditional criminogenic risk factors as treatment targets could further our understanding of the application of the risk, need and responsivity principles at the level of the individual (as opposed to groups of offenders) and specifically for female offenders. A similar study was conducted examining matching the court-ordered services with treatment needs of young offenders (Vieira, Skilling, & Peterson-Badali, 2009) and found that a better match between assessed and treated needs improved the prediction of recidivism over that of risk score alone. They also found that a better match between assessed responsivity factors and addressed responsivity factors predicted lower recidivism outcome, but did not improve the prediction of recidivism over and above risk and matched needs. As this study included only 19

girls, a follow up study by Vitopoulos (2011) examined the impact of gender on matching assessed and treated needs to understand the RNR model at the individual level for a small sample of young offenders. She found matching services to RNR needs was more potent for males than females. Thus, such an exploration may be especially important for adult female offenders. Given that the risk/need areas identified in both the current study and those listed above on youth predicted outcome effectively, this potential gender difference in the application of the RNR principles may point to differences in treatment response, which could underlie treatment delivery or dosage concerns. Further, the examination of responsivity as described above has focused on the general responsivity principle (Vieira, Skilling, & Peterson-Badali, 2009; Vitopoulos, 2011). Examining factors like the gender informed variables from the current study could address the specific responsivity principle and may also explain the potential gender difference found in youth.

Finally, while these youths' needs appeared to be assessed prior to the court directing treatment, risk assessments can occur at different times throughout an offender's contact with the law. If risk assessments were conducted for offenders while in custody, as in the current study, it is possible that some interventions could overlap with the assessment period and thus, such a study would need to consider (and could perhaps clarify) when the impact of treatment is realized. While risk assessment instruments, such as the LSI-R, LS/CMI and VRS included in the current study, contain ratings of dynamic needs which allow for the possibility of change in response to treatment, the rating guides suggest that some time is needed to demonstrate behavioural change after the completion of an intervention to ensure that risk is indeed reduced before lowering the offender's rating on a particular item or domain. This would be an important methodological consideration for future research, as these risk assessment instruments scores may not reflect early treatment changes, but looking at the impact of treatment only following assessment could mean that the full impact of treatment is not captured in either the risk/need rating or the measurement of treatment. This could be an important contribution to our growing understanding of treatment dosage as well.

4.1.7. Conclusion.

The current study provided support for the utility of the LSI-R, LS/CMI and VRS for federal female offenders. It also highlighted the likelihood that the predictive utility of these instruments could be improved for federal female offenders by considering gender informed

variables, especially in combination. While the present data did not clearly identify which gender informed composites of substance abuse and abuse history were most useful in improving the prediction of a variety of recidivism outcomes, childhood emotional abuse did appear the most promising. Moreover, by reaffirming the complex interplay of these gender informed variables in contributing to reoffending, it seems that exploring more proximal and dynamic constructs related to the often historic gender informed variables explored to date (e.g., abuse history) may be the next step in better assessing women offenders' risk and identifying treatment targets for desistence.

Finally, using risk assessment instruments is only one (albeit important) part of conducting a useful forensic assessment. A number of other factors should be considered in addition to one's risk score. These factors will vary depending upon the context and purpose the forensic assessment, but would include consideration of responsivity factors like cultural influences, reading and learning ability that could affect treatment efficacy, health and mental health concerns that could impact overall wellbeing, as well as unique individual strengths like having concrete or emotional supports available, and additional motivating factors that could impact crime desistance. In fact, recent work by Van Voorhis, Wright, Salisbury, and Bauman (2010) has highlighted that the gender informed factors that should be considered when conducting forensic assessments with women may very well differ depending upon what is being predicted (e.g., different factors predicting institutional adjustment versus successful community reintegration). While work toward improving risk assessment instruments may continue to include explorations of some of these factors, ongoing recognition that some factors and the unique combination of these factors will be *woman*-specific requires that we continue also to focus on the *process* of forensic assessments which complement and round out the use of a valid risk assessment tool. A focus on the process allows for a more holistic understanding of an individual woman's needs while also allowing assessing clinicians to build upon the utility of empirically supported commonalities among offenders to form a more integrated case conceptualization of the individual. While researchers have given the attention to the importance of understanding women offenders more holistically, this process is likely important for all offenders (in fact, some gender informed variables, like childhood abuse are found to be relevant for men as well, for example see Topitzes, Mersky, & Reynolds, 2011). This conceptualization moves beyond focusing on modifying risk scores/levels (e.g., with clinical overrides, or

additional risk factors) to recognizing needs and resources more generally in context to provide a more complete and useful assessment. Moreover, it will likely include focusing on addressing concerns that are relevant and prevalent for offenders, albeit not directly related to reoffending (Blanchette & Brown, 2006; Andrews et al., 2011b) The additional sections of the LS/CMI focusing on case management (only section 1, the general risk/need factors, was included in this study) and the scorings of strengths in the general risk/need section bring some concrete focus to the importance of considering other factors beyond the risk assessment instrument scores to provide improved service to the client and community on whole. Although such a process should (and often may), be used with the LSI-R and VRS as well if they are the risk assessment instrument selected for use in a risk assessment. Future research should focus not only on the utility and refinement of risk assessment instruments, but on improving our assessment processes on whole, recognizing risk assessment instruments are but one, albeit important, part of a thorough and useful assessment.

LIST OF APPENDICES

APPENDICIES

5.1 Appendix A: Frequency Distribution of LSI-R Total Scores

Table A1: Frequency Distribution of LSI-R Total Scores

LSI-R Total Score	Frequency	Percent	Cumulative Percent
0	1	1.0	1.0
7	2	2.0	3.0
8	3	3.0	5.9
10	1	1.0	6.9
11	1	1.0	7.9
12	2	2.0	9.9
13	2	2.0	11.9
14	1	1.0	12.9
15	3	3.0	15.8
16	3	3.0	18.8
17	3	3.0	21.8
18	1	1.0	22.8
20	4	4.0	26.7
21	2	2.0	28.7
22	2	2.0	30.7
23	2	2.0	32.7
24	2	2.0	34.7
25	4	4.0	38.6
26	1	1.0	39.6
28	2	2.0	41.6
29	6	5.9	47.5
30	5	5.0	52.5
31	2	2.0	54.5
32	4	4.0	58.4
33	4	4.0	62.4
34	2	2.0	64.4
35	7	6.9	71.3
36	4	4.0	75.2
37	4	4.0	79.2
39	2	2.0	81.2
40	5	5.0	86.1

Table A1: Frequency Distribution of LSI-R Total Scores Continued

LSI-R Total Score	Frequency	Percent	Cumulative Percent
41	5	5.0	91.1
42	3	3.0	94.1
43	1	1.0	95.0
44	1	1.0	96.0
45	3	3.0	99.0
48	1	1.0	100.0
Total	101	100.0	

5.2 Appendix B: Frequency Distribution of LS/CMI Total Scores

Table A2: Frequency Distribution of LS/CMI Total Scores

LS/CMI Total Score	Frequency	Percent	Cumulative Percent
6	2	2.0	2.0
7	5	5.0	6.9
8	1	1.0	7.9
9	4	4.0	11.9
10	2	2.0	13.9
11	2	2.0	15.8
12	1	1.0	16.8
13	2	2.0	18.8
14	1	1.0	19.8
15	4	4.0	23.8
16	4	4.0	27.7
17	4	4.0	31.7
18	1	1.0	32.7
19	1	1.0	33.7
20	8	7.9	41.6
21	8	7.9	49.5
22	2	2.0	51.5
23	3	3.0	54.5
24	4	4.0	58.4
25	4	4.0	62.4
26	4	4.0	66.3
27	2	2.0	68.3
28	1	1.0	69.3
29	3	3.0	72.3
30	3	3.0	75.2
31	3	3.0	78.2
32	2	2.0	80.2
33	6	5.9	86.1
34	3	3.0	89.1
35	2	2.0	91.1
36	3	3.0	94.1
37	4	4.0	98.0
38	1	1.0	99.0
39	1	1.0	100.0
Total	101	100.0	

5.3 Appendix C: Frequency Distribution of VRS Total Scores

Table A3: Frequency Distribution of VRS Total Scores

VRS Grand Total	Frequency	Percent	Cumulative Percent
3.00	2	2.0	2.0
4.00	1	1.0	3.0
4.16	1	1.0	4.0
5.00	1	1.0	5.0
5.20	3	3.0	7.9
5.42	1	1.0	8.9
8.00	4	4.0	12.9
8.32	2	2.0	14.9
9.00	3	3.0	17.8
9.36	1	1.0	18.8
10.00	1	1.0	19.8
10.40	2	2.0	21.8
12.00	4	4.0	25.7
13.00	2	2.0	27.7
14.00	1	1.0	28.7
14.08	1	1.0	29.7
14.08	1	1.0	30.7
14.56	1	1.0	31.7
15.00	1	1.0	32.7
15.60	1	1.0	33.7
16.00	3	3.0	36.6
16.64	1	1.0	37.6
16.96	1	1.0	38.6
19.00	3	3.0	41.6
19.76	1	1.0	42.6
20.00	1	1.0	43.6
21.00	2	2.0	45.5
22.00	2	2.0	47.5
23.00	1	1.0	48.5
24.00	2	2.0	50.5
25.00	2	2.0	52.5
26.00	2	2.0	54.5
27.00	1	1.0	55.4
28.08	1	1.0	56.4

Table A3: Frequency Distribution of VRS Total Scores Continued

VRS Grand Total	Frequency	Percent	Cumulative Percent
29.00	5	5.0	61.4
31.00	1	1.0	62.4
32.00	4	4.0	66.3
34.00	4	4.0	70.3
34.32	1	1.0	71.3
35.00	1	1.0	72.3
35.36	2	2.0	74.3
36.00	3	3.0	77.2
37.00	2	2.0	79.2
38.00	1	1.0	80.2
39.00	4	4.0	84.2
40.00	1	1.0	85.1
41.00	1	1.0	86.1
42.00	2	2.0	88.1
43.00	2	2.0	90.1
43.16	1	1.0	91.1
43.68	1	1.0	92.1
47.00	1	1.0	93.1
51.00	2	2.0	95.0
56.00	1	1.0	96.0
58.00	1	1.0	97.0
59.00	1	1.0	98.0
61.36	1	1.0	99.0
63.00	1	1.0	100.0
Total	101	100.0	

5.4 Appendix D: Frequency Distribution of SIR-R1 Total Scores

Table A4: Frequency Distribution of SIR-R1 Total Scores

SIR-R1 Total Score	Frequency	Percent	Cumulative Percent
-17	1	1.0	1.0
-15	1	1.0	2.0
-11	2	2.0	4.0
-10	1	1.0	5.0
-8	1	1.0	5.9
-7	6	5.9	11.9
-6	1	1.0	12.9
-5	1	1.0	13.9
-4	1	1.0	14.9
-3	2	2.0	16.8
-2	4	4.0	20.8
-1	6	5.9	26.7
0	3	3.0	29.7
1	3	3.0	32.7
2	6	5.9	38.6
3	4	4.0	42.6
4	2	2.0	44.6
5	3	3.0	47.5
6	5	5.0	52.5
7	5	5.0	57.4
8	2	2.0	59.4
9	3	3.0	62.4
10	6	5.9	68.3
11	5	5.0	73.3
12	5	5.0	78.2
13	5	5.0	83.2
14	4	4.0	87.1
15	5	5.0	92.1
16	1	1.0	93.1
17	6	5.9	99.0
19	1	1.0	100.0
Total	101	100.0	

5.5.1 Appendix E: LSI-R Item Interrater Reliability, Items 1 to 30.

Table A5: LSI-R Item Interrater Reliability, Items 1 to 30.

Item	Interrater reliability (ICC)
1. LSI-R: Prior adult convictions?	1.00
2. LSI-R: Two or more prior adult convictions?	.875
3. LSI-R: Three or more prior adult convictions?	.873
4. LSI-R: Three or more present offenses?	1.00
5. LSI-R: Arrested under age 16?	1.00
6. LSI-R: Ever incarcerated upon conviction?	1.00
7. LSI-R: Escape history from a correctional facility?	1.00
8. LSI-R: Ever punished for institutional misconduct?	1.00
9. LSI-R: Charge laid or probation/parole suspended during prior community supervision?	1.00
10. LSI-R: Official record of assault/violence?	1.00
11. LSI-R: Currently unemployed?	1.00
12. LSI-R: Frequently unemployed?	.825
13. LSI-R: Never employed for a full year?	.588
14. LSI-R: Ever fired?	.353
15. LSI-R: Less than regular grade 10?	.873
16. LSI-R: Less than regular grade 12?	.750
17. LSI-R: Suspended or expelled at least once?	.133 *
18. LSI-R: Participation/performance	.986
19. LSI-R: Peer interactions	.940
20. LSI-R: Authority interactions	.986
21. LSI-R: Financial problems	.576
22. LSI-R: Reliance upon social assistance	.364
23. LSI-R: Dissatisfaction with marital or equivalent situation	.690
24. LSI-R: Non-rewarding, parental	.843
25. LSI-R: Non-rewarding, other relatives	.690
26. LSI-R: Criminal-Family/spouse	.674
27. LSI-R: Unsatisfactory	.578
28. LSI-R: 3 or more address changes last year	.283
29. LSI-R: High crime neighbourhood	.759
30. LSI-R: Absence of organized activity	.488

* Note, this low values reflects a scoring difference whereby one rater consistently felt there was inadequate information to make a rating and the other rated the items as absent. The hit rates were an exact match.

5.5.2 Appendix F: LSI-R Item Interrater Reliability, Items 31 to 54.

Table A6: LSI-R Item Interrater Reliability, Items 31 to 54.

Item	Interrater reliability (ICC)
31. LSI-R: Could make better use of time	.762
32. LSI-R: A social isolate	.774
33. LSI-R: Some criminal acquaintances	.650
34. LSI-R: Some criminal friends	.682
35. LSI-R: Few anti-criminal acquaintances	.618
36. LSI-R: Few anti-criminal friends	.562
37. LSI-R: Alcohol problem, ever	1.00
38. LSI-R: Drug problem, ever	1.00
39. LSI-R: Alcohol problem, currently	.892
40. LSI-R: Drug problem, currently	.970
41. LSI-R: Law violations	.750
42. LSI-R: Marital/Family	.750
43. LSI-R: School/work	.462
44. LSI-R: Medical	.604
45. LSI-R: Other indicators	.875
46. LSI-R: Moderate interference	.462
47. LSI-R: Severe interference, active psychosis	1.00
48. LSI-R: Mental health treatment, past	.656
49. LSI-R: Mental health treatment, present	.200
50. LSI-R: Psychological assessment indicated	.375
51. LSI-R: Supportive of crime	.728
52. LSI-R: Unfavourable toward convention	.678
53. LSI-R: Poor, toward sentence	.543
54. LSI-R: Poor, toward supervision	.440

5.6.1 Appendix G: LS/CMI Item Interrater Reliability, Items 1 to 30.

Table A7: LS/CMI Item Interrater Reliability, Items 1 to 30.

Item	Interrater reliability (ICC)
1. LS/CMI: Prior youth dispositions or adult convictions?	1.00
2. LS/CMI: Two or more prior youth/adult dispositions/convictions?	1.00
3. LS/CMI: Three or more prior youth/adult dispositions/convictions?	1.00
4. LS/CMI: Three or more present offenses?	1.00
5. LS/CMI: Arrested or charged under age 16?	1.00
6. LS/CMI: Ever incarcerated upon conviction?	1.00
7. LS/CMI: Ever punished for institutional misconduct or a behavior report?	1.00
8. LS/CMI: Charge laid, probation breached, or parole suspended during prior community supervision?	1.00
9. LS/CMI: Currently unemployed?	1.00
10. LS/CMI: Frequently unemployed?	.488
11. LS/CMI: Never employed for a full year?	.600
12. LS/CMI: Less than regular grade 10 or equivalent?	.873
13. LS/CMI: Less than regular grade 12 or equivalent?	.875
14. LS/CMI: Suspended or expelled least once?	.133*
15. LS/CMI: Participation/performance	.982
16. LS/CMI: Peer interactions	.926
17. LS/CMI: Authority interactions	.981
18. LS/CMI: Dissatisfaction with marital or equivalent situation	.752
19. LS/CMI: Non-rewarding, parental	.870
20. LS/CMI: Non-rewarding, other relatives	.815
21. LS/CMI: Criminal-Family/spouse	.674
22. LS/CMI: Absence of recent participation in organized activity	.462
23. LS/CMI: Could make better use of time	.578
24. LS/CMI: Some criminal acquaintances	1.00
25. LS/CMI: Some criminal friends	.812
26. LS/CMI: Few anti-criminal acquaintances	.611
27. LS/CMI: Few anti-criminal friends	.679
28. LS/CMI: Alcohol problem, ever	1.00
29. LS/CMI: Drug problem, ever	1.00
30. LS/CMI: Alcohol problem, currently	.908

* Note, this low values reflects a scoring difference whereby one rater consistently felt there was inadequate information to make a rating and the other rated the items as absent. The hit rates were an exact match.

5.6.2 Appendix H: LS/CMI Item Interrater Reliability, Items 31 to 43.

Table A8: LS/CMI Item Interrater Reliability, Items 31 to 43.

Item	Interrater reliability (ICC)
31. LS/CMI: Drug problem, currently	.920
32. LS/CMI: Law violations	.632
33. LS/CMI: Marital/Family	.517
34. LS/CMI: School/work	.462
35. LS/CMI: Medical or other clinical indicators	.750
36. LS/CMI: Supportive of crime	.728
37. LS/CMI: Unfavorable toward convention	.678
38. LS/CMI: Poor, toward sentence/ offence	.333
39. LS/CMI: Poor, toward supervision/treatment	.440
40. LS/CMI: Specialized assessment for antisocial pattern	.600
41. LS/CMI: Early and diverse antisocial behavior	.865
42. LS/CMI: Criminal attitude	.543
43. LS/CMI: Pattern of generalized trouble	.741

5.7 Appendix I: VRS Item Interrater Reliability.

Table A9: VRS Item Interrater Reliability.

Item	Interrater reliability (ICC)
s1VRS: Current age	1.00
s2VRS: Age at first violent conviction	.623
s3VRS: Number of juvenile convictions	.873
s4VRS: Violence through lifespan	.951
s5VRS: Prior release failures/escapes	.851
s6VRS: Stability of family upbringing	1.00
d1VRS: violent lifestyle	1.00
d2VRS: Criminal personality	.959
d3VRS: Criminal attitudes	.686
d4VRS: Work ethic	.333
d5VRS: Criminal peers	.329
d6VRS: Interpersonal aggression	.864
d7VRS: Emotional control	.817
d8VRS: Violence during incarceration	.361
d9VRS: Weapon use	.920
d10VRS: Insight into violence	.590
d11VRS: Mental illness	.885
d12VRS: Substance abuse	.967
d13VRS: Stability of relationships	.801
d14VRS: Community Support	.925
d15VRS: Released to high risk situations	.728
d16VRS: Violence cycle	.660
d17VRS: Impulsivity	.742
d18VRS: Cognitive distortion	.769
d19VRS: Compliance with supervision	.800
d20VRS: Security level of release institution	.554

5.8 Appendix J: SIR-R1 Item Interrater Reliability.

Table A10: SIR-R1 Item Interrater Reliability.

Item	Interrater reliability (ICC)
1. Current offence	.881
2. Age at admission	1.00
3. Previous incarceration	.982
4. Revocation or forfeiture	.714
5. Act of escape	.462
6. Security classification	1.00
7. Age at first adult conviction	1.00
8. Previous convictions for assault	1.00
9. Marital status at most recent admission	.865
10. Interval at risk since last offence	.979
11. Number of dependants at most recent admission	1.00
12. Current total aggregate sentence	1.00
13. Previous convictions for sexual offence(s)	1.00
14. Previous convictions for break and enter	.744
15. Employment status at arrest	.825

5.9 Appendix K: Correlations between Outcome Variables

Table A 11: Correlations between Outcome Variables

Outcome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22 ***
2	.451**	1																				
3	.488**	.950**	1																			
4	.598**	.605**	.581**	1																		
5	.741**	.507**	.539**	.591**	1																	
6	.444**	.724**	.728**	.604**	.599**	1																
7	.448**	.725**	.829**	.543**	.604**	.912**	1															
8	.464**	.462**	.460**	.775**	.625**	.717**	.607**	1														
9	.276**	.174	.232*	.289**	.444**	.233*	.291**	.266**	1													
10	.290**	.105	.145	.285**	.464**	.171	.228*	.242*	.869**	1												
11	.466**	.342**	.369**	.491**	.417**	.274**	.321**	.368**	.408**	.421**	1											
12	.235*	.208*	.195	.234*	.259**	.118	.152	.100	.301**	.354**	.505**	1										
13	.471**	.319**	.335**	.475**	.410**	.271**	.306**	.389**	.366**	.377**	.961**	.519**	1									
14	.234*	.196	.177	.238*	.251*	.108	.135	.104	.281**	.337**	.497**	.995**	.518**	1								
15	.179	.357**	.344**	.172	.221*	.189	.241*	.047	.304**	.364**	.339**	.501**	.290**	.442**	1							
16	-.207	-.208	-.199	-.078	-.166	-.144	-.146	.068	-.094	-.062	.a	-.232	.147	-.210	-.190	1						
17	-.250	-.219	-.207	-.103	-.184	-.157	-.153	.040	-.070	-.017	.a	-.265	.a	-.248	-.149	.998**	1					
18	-.037	-.329	-.277	-.371	.122	-.219	-.142	-.046	.190	.048	.a	.248	.259	.230	.a	.860**	.833** (n=10)	1				
19	.290**	.202*	.180	.250*	.318**	.150	.143	.192	.295**	.378**	.386**	.545**	.401**	.548**	.319**	-.242	-.255	.091	1			
20	.256**	.364**	.334**	.289**	.301**	.237*	.256**	.171	.281**	.337**	.384**	.765**	.399**	.769**	.411**	-.271	-.292*	.073	.867**	1		
21	.270**	.239*	.254*	.286**	.268**	.125	.165	.172	.241*	.235*	.556**	.713**	.565**	.695**	.422**	-.025	-.058	.250	.308**	.430**	1	
22***	.447**	.497**	.453**	.460**	.404**	.421**	.377**	.312**	.423**	.466**	.924**	.918**	.896**	.878**	.555**	-.218	-.194	-.083	.770**	.770**	.702**	1
23***	-.460**	-.493**	-.440**	-.483**	-.391**	-.408**	-.351**	-.340**	-.392**	-.417**	-.923**	-.942**	-.916**	-.928**	-.362**	.299*	.370**	-.138	-.774**	-.778**	-.755**	-.929**

N = 101 except where indicated; *** Spearman correlations for ranked data, remaining are Pearson correlations; ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed); a. Cannot be computed because at least one of the variables is constant; No. = Number; 1 = Presence of Misconduct during Index, 2 = No. of Misconduct Charges during Index, 3 = No. of Minor Misconduct Charges during Index, 4 = Presence of Serious Misconduct Charges during Index, 5 = Presence of Misconduct After Assessment Only, 6 = No. of Misconduct Charges After Assessment Only, 7 = No. of Minor Misconduct After Assessment Only, 8 = Presence of Serious Misconduct Charges After Assessment Only, 9 = Presence of Revocations, 10 = No. of Revocations, 11 = Recidivism Present, 12 = No. Reconvictions, 13 = Nonviolent Reconviction Present, 14 = No. Nonviolent Reconvictions, 15 = Violent Reconvictions Present, 16 = Time to 1st Reconviction (n = 52), 17 = Time to 1st Nonviolent Reconviction (n = 50), 18 = Time to 1st Violent Reconviction (n = 11), 19 = Maximum Time Sentenced for Reconvictions, 20 = Sum of Aggregate Incarceration Sentences, 21 = Sum of Aggregate Probation Sentences, 22 = Maximum of Severity Index score, 23 = Minimum of Offence Categories Ranked by Seriousness (OCRS) Score

5.10 Appendix L: Chi Squares for General Recidivism Rates by Risk Level.

Table A12: Chi Squares for General Recidivism Rates by Risk Level.

Risk Instrument	Risk Levels	Recidivism		Of Total	Chi Square
		No	Yes	Sample	(df)
LSI-R	Low	11 (91.7%)	1 (8.3%)	12 (11.9%)	
	low/moderate	15 (71.4%)	6 (28.6%)	21(20.8%)	
	moderate	12 (40.0%)	18 (60.0%)	30 (29.7%)	
	medium/high	6 (25.0%)	18 (75.0%)	24 (23.8%)	
	High	5 (35.7%)	9 (64.3%)	14(13.9%)	20.462 (4) **
LS/CMI	Low	14 (100.0%)	0 (.0%)	14 (13.9%)	
	medium	15 (75.0%)	5 (25.0%)	20 (19.8%)	
	High	13 (33.3%)	26 (66.7%)	39 (38.6%)	
	very high	7 (25.0%)	21 (75.0%)	28 (27.7%)	30.271 (3) **
VRS	Low	19 (73.1%)	7 (26.9%)	26 (25.7%)	
	medium	27 (51.9%)	25 (48.1%)	52 (51.5%)	
	High	3 (13.0%)	20 (87.0%)	23 (22.8%)	18.107 (2) **
SIR-R1	low risk	38 (71.7%)	15 (28.3%)	53 (52.5%)	
	low to moderate	5 (27.8%)	13 (72.2%)	18 (17.8%)	
	moderate	5 (31.3%)	11 (68.8%)	16 (15.8%)	
	moderate to high	0 (.0%)	9 (100.0%)	9 (8.9%)	
	High	1 (20.0%)	4 (80.0%)	5 (5.0%)	26.521 (4)**
Total		49 (48.5%)	52 (51.5%)	101 (100.0%)	

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

5.11 Appendix M: Chi Squares for Violent Recidivism Rates by Risk Level.

Table A13: Chi Squares for Violent Recidivism Rates by Risk Level.

Risk Instrument	Risk Levels	Violent Recidivism		Of Total	Chi Square
		No	Yes	Sample	(df)
LSI-R	low	12 (100.0%)	0 (.0%)	12 (11.9)	
	low/moderate	21 (100%)	0 (.0%)	21 (20.8)	
	moderate	28 (93.3%)	2 (6.7%)	30 (29.7)	
	medium/high	19 (79.2%)	5 (20.8%)	24 (23.8)	
	high	10 (71.4%)	4 (28.6%)	14 (13.9)	11.539 (4)*
LS/CMI	low	14 (100.0%)	0 (.0%)	14 (13.9%)	
	medium	20 (100.0%)	0 (.0%)	20 (19.8%)	
	high	35 (89.7%)	4 (10.3%)	39 (38.6%)	
	very high	21 (75.0%)	7 (25.0%)	28 (27.7%)	9.915 (3) *
VRS	low	26 (100.0%)	0 (.0%)	26 (25.7%)	
	medium	48 (92.3%)	4 (7.7%)	52 (51.5%)	
	high	16 (69.6%)	7 (30.4%)	23 (22.8%)	12.778 (2) **
SIR-R1	low risk	50 (94.3%)	3 (5.7%)	53 (52.5%)	
	low to moderate	18 (100.0%)	0 (.0%)	18 (17.8%)	
	moderate	11 (68.8%)	5 (31.3%)	16 (15.8%)	
	moderate to high	7 (77.8%)	2 (22.2%)	9 (8.9%)	
	high	4 (80.0%)	1 (20.0%)	5 (5.0%)	12.146 (4)*
Total		90 (89.1%)	11 (10.9%)		

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

5.12 Appendix N: Examination of New SIR-R1 Cutoff Scores Based on Total Score Frequencies.

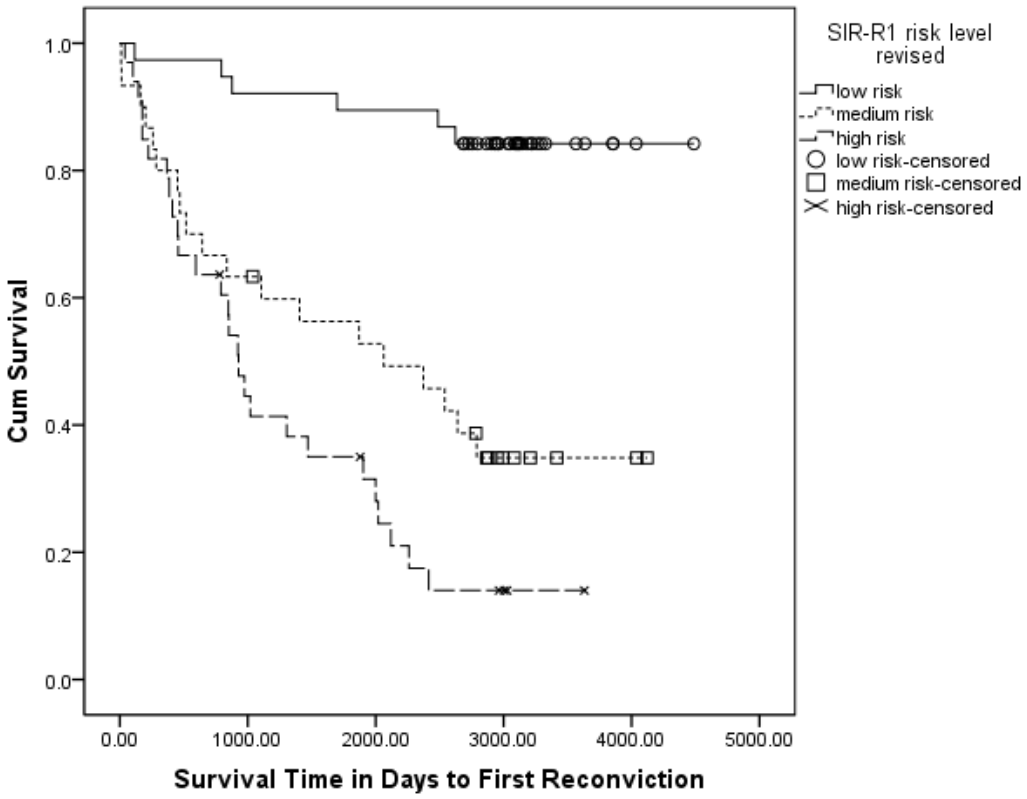

Figure A1: Survival Analysis: Cumulative General Recidivism Failure Rate as a Function of SIR-R1 Revised Risk Levels.

Table A14: Mean Survival Time for General Recidivism for each Revised SIR-R1 Risk Level.

SIR-R1	Mean[a] Survival Time			
			95% Confidence Interval	
Risk Levels Revised	Estimate	Std. Error	Lower Bound	Upper Bound
low risk	4004.711	190.105	3632.105	4377.317
medium risk	2164.170	300.161	1575.855	2752.486
high risk	1350.360	204.068	950.388	1750.333

a. Estimation is limited to the largest survival time if it is censored.
Wilcoxon (2) = 32.848, p < .001

REFERENCES

Allen, J. S., Rupert, V., Spatafora, K., Windell, J.O., Gaulier, B., & Conti, K.L. (2002). Differentiating violent from non-violent female offenders using the Jesness inventory. *Personality and Individual Differences, 35,* 101-108.

Andrews, D. A. (2000). Principles of effective correctional programming. In L. L. Motiuk & R. C. Serin (Eds.), Compendium 2000 on effective correctional programming (pp. 9 – 17). Ottawa, ON: Correctional Service Canada.

Andrews, D. A., & Bonta, J. L. (1995). *Level of Service Inventory-Revised manual*. Toronto, Ontario, Canada: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct* (3rd ed.). Cincinnati, Ohio: Anderson Publishing Company.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *Manual for the Level of Service/Case Management Inventory (LS/CMI).* Toronto, Canada: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2011). The risk-need-responsivity (RNR) model: Does adding the Good Lives Model contribute to effective crime prevention? *Criminal Justice and Behavior, 38(7),* 735 – 755.

Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011a). Sources of variability in estimates of predictive validity: A specification with Level of Service general risk and need. *Criminal Justice and Behavior, 38(5) ,* 413 – 432.

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., et al., (2011b). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology, XX,* XX – XX.

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology, 28 (3),* 369- 404.

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior, 28,* 490-521.

Bartko, J. J. (1991). Measure and reliability: Statistical thinking considerations. *Schizophrenia Bulletin, 17,* 483- 489.

Belanger, B. (2001). Sentencing in adult criminal courts, 1999/00. *Juristat, 21 (10),* 1 - 20.

Belcourt, R., Nouwens, T., & Lefebvre, L.(1993). Examining the unexamined: Recidivism among female offenders. *Forum on Corrections Research, 5(3),* 10 – 14.

Benda, B. (2005).Gender Differences in Life-Course Theory of Recidivism: A Survival Analysis. *International Journal of Offender Therapy and Comparative Criminology, 49,* 325 – 342.

Berzenski, S. R., & Yates, T. M. (2010). A developmental process analysis of the contribution of childhood emotional abuse to relationship violence. *Journal of Aggression, Maltreatment and Trauma, 19*(2), 180–203.

Blanchette, K. (1997a). Classifying federal female offenders for correctional interventions. *Forum on Corrections Research, 9(1),* 36-41.

Blanchette, K. (1997b). Comparing violent and non-violent female offenders on risk and need. *Forum on Corrections Research, 9(2),* 14-18.

Blanchette, K. (2000). Chapter 20: Effective correctional practice with women offenders. In L. L. Motiuk & R. C. Serin (Eds.), *Compendium 2000 on effective correctional programming* (pp. 160 - 173). Ottawa, ON: Correctional Service Canada.

Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of women offenders.* West Sussex, England: John Wiley & Sons Ltd.

Blanchette, K., & Motiuk, L. (2002). Classifying female offenders for effective intervention: Application of the case-based principles of risk and need. *Forum on Corrections Research, 14(1),* 31 - 35.

Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). The prediction of recidivism among federally sentenced offenders: A re-validation of the SIR scale. *Canadian Journal of Criminology, 1996,* 61 - 79.

Bonta, J., Lipinski, S., & Martin, M. (1993). Recidivists tend to be …, *Forum on Corrections Research, 5(3),* 14 - 16.

Bonta, J., Pang, B., & Wallace-Capretta, S. (1995). Predictors of recidivism among incarcerated female offenders. *The Prison Journal, 75 (3),* 277 - 294.

Bonta, J., Rugge, T., & Dauvergne, M. (2003). The reconviction rate of federal offenders. Ottawa, ON: Solicitor General Canada.

Bloom, B., Owen, B., & Covington, S. (2003). *Gender-Responsive Strategies: Research, Practice, and Guiding Principles for Women Offenders.* Washington, D.C.: U.S.: Department of Justice, National Institute of Corrections.

Brewster, D. R., & Sharp, S.F. (2002). Educational programs and recidivism in Oklahoma: Another look. *The Prison Journal, 82(3),* 314-334.

Brown, S. L. & Motiuk, L. L. (2005). *The dynamic factors identification and analysis (DFIA) component of the offender intake assessment (OIA) process: A meta-analytic, psychometric and consultative review.* (R-164) Ottawa, ON: Research Branch, Correctional Service of Canada.

Cabin, R. J. & Mitchell, R. J. (2000). To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America, 81,* 246 – 248.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36(6),* 567 - 590.

Carlson, B. E. & Shafer, M. S. (2010). Traumatic histories and stressful life events of incarcerated parents: Childhood and adult trauma histories. *The Prison Journal, 90(4),* 475 – 493.

Caspi, A., Lynam, D., Moffitt, T. E., & Silva, P. A. (1993). Unraveling girls' delinquency: Biological, dispositional, and contextual contributions to adolescent misbehavior. *Developmental Psychology, 29 (1),* 19-30.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Collins, R. E. (2010). The effect of gender on violent and nonviolent recidivism: A meta-analysis. *Journal of Criminal Justice, 38,* 675 – 684.

Cormier, R. B. (1997). Yes, SIR! A stable risk prediction instrument. *Forum on Corrections Research*, 9(1), Ottawa, ON: Correctional Service of Canada.

Correctional Service Canada (1994). *Correctional program strategy for federally sentenced women.* Ottawa, ON: Federally Sentenced Women Program.

Correctional Service Canada (2003). *Offender intake assessment and correctional planning: Standard operating practices (SOPs)* (Publication No.700-04). Ottawa, ON: McClung, L.

Correctional Service Canada (2005). *Basic facts about the Correctional Service of Canada.* Ottawa, ON: Public Works and Government Services Canada.

Cortina. J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104.

Coulson, G., Ilacqua, G., Nutbrown, V., Giulekas, D., & Cudjoe, F. (1996). Predictive utility of the LSI for incarcerated female offenders. *Criminal Justice and Behavior*, *23*, 427-439.

Claussen, A. H., & Crittenden, P. M. (1991). Physical and psychological maltreatment: relations among types of maltreatment. *Child Abuse & Neglect, 15 (1),* 5 – 18.

Daly, K. (1992). Women's pathways to felony court: feminist theories of lawbreaking and problems of representation. *Southern California Review of Law and Women's Studies, 2,* 11 – 52.

Dell, C. A. & Boe, R. (2000). *An examination of aboriginal and Caucasian women offender risk and needs factors.* Ottawa, ON: Research Branch, Correctional Service of Canada.

de Vogel, V., & de Ruiter, C. (2005). The HCR-20 in personality disordered female offenders: a comparison with a matched sample of males. *Clinical Psychology and Psychotherapy, 12*, 226-240.

Dolan, M. & Fullam, R. (2007). The validity of the Violence Risk Scale second edition (VRS-2) in a British forensic inpatient sample. *The Journal of Forensic Psychiatry & Psychology, 18(3),* 381 – 393.

Dolan, M., Fullam, R., Logan, C., & Davies, G. (2008). The Violence Risk Scale Second Edition (VRS-2) as a predictor of institutional violence in a British forensic inpatient sample. *Psychiatry Research, 158,* 55 – 65.

Dowden, C. & Andrews, D. A. (1999). What works for female offenders: A meta-analytic review. *Crime and Delinquency*, *45*, 438-452.

Dowden, C., & Blanchette, K. (2002). An evaluation of the effectiveness of substance abuse programming for female offenders. *International Journal of Offender Therapy and Comparative Criminology, 46(2),* 220-230.

Dowden, C., & Brown, S. L. (2002). The role of substance abuse factors in predicting recidivism: A meta-analysis. *Psychology, Crime, & Law, 8*, 243- 264.

Federally Sentenced Women Program (1994). *Correctional Program Strategy for Federally Sentenced Women* (FSW N-18). Ottawa, ON: Correctional Service of Canada.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Folsom, J., & Atkinson, J. (2007). The generalizability of the LSI-R and the CAT to the prediction of recidivism in female offenders. *Criminal Justice and Behavior, 34,* 1044-1056.

Fortin, D. (2004). *Program strategy for women offenders.* Ottawa, ON: Programs for Women Offenders.

Funk, S. J. (1999). Risk assessment for juveniles on probation: A focus on gender. *Criminal Justice and Behavior, 26,* 44-68.

Garamszegi, L. Z. (2006). Comparing effect sizes across variables: generalization without need for Bonferroni correction. *Behavioral Ecology, 17 (4),* 682 -687.

Gendreau, P., Goggin, C. E., & Law, M. A. (1997). Predicting prison misconducts. *Criminal Justice and Behavior, 24,* 414-431.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior, 29,* 397-426.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34,* 575-607.

Giordano, P. C. & Cernkovich, S. A. (1997). Chapter 46: Gender and antisocial behavior. In D. M. Stoff, J. Breiling & J. D. Maser (Eds.), *Handbook of antisocial behavior* (pp.496 – 510). New York: John Wiley & Sons Inc.

Girard, L. & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory – Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior, 31(2),* 150 – 181.

Gobeil, R. & Robeson Barrett, M. (2007). *Rates of recidivism for women offenders.* (Publication Number R-192).Ottawa, ON: Correctional Service Canada.

Gordon, A. (1998). *The reliability and validity of the Violence Risk Scale – Experimental Version 1.* Unpublished masters thesis, University of Saskatchewan, SK., Canada.

Gover, A. R., Perez, D. M., & Jennings, W. G. (2008). Gender differences in factors contributing to institutional misconduct. *The Prison Journal, 88(3),* 378 – 403.

Green, S. B., Lissitz, R. W, & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37,* 827-838.

Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: can structured clinical assessments predict the short term risk of inpatient violence? *The Journal of Forensic Psychiatry & Psychology, 15(2),* 278-292.

Hannah- Moffat, K. & Shaw, M. (2001). *'Taking Risks' – Gender, diversity, security classification and risk/needs assessment with Canadian federally sentenced women.* Ottawa, ON: Status on Women Canada.

Hannah-Moffat, K. (2009). Gridlock or mutability: reconsidering "gender" asnd risk assessment. *Criminology and Public Policy, 8(1),* 209 – 219.

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143,* 29 – 36.

Hardyman, P. L. & Van Voorhis, P. (2004). Developing gender-specific classification systems for women offenders. Washington, DC: USDOJ, National Institute of Corrections.

Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist- Revised.* Toronto, Ontario, Canada: Multi-Health Systems.

Hastings, M. E., Krishnan, S., Tangney, J. P., & Stuewig, J. (2011). Predictive and incremental validity of the Violence Risk Appraisal Guide scores with male and female jail inmates. *Psychological Assessment, 23,* 174 – 183.

Hatch, A., & Faith, K. (1989). Female offender in Canada: A statistical profile. *Canadian Journal of Women and the Law, 3(2),* 423-456.

Heilbrun, K., Dematteo, D., Fretz, R., Erickson, J., Yashuhara, K., & Anumba, N. (2008). How "specific" are gender-specific rehabilitation needs? An empirical analysis. *Criminal Justice and Behavior, 35 (11),* 1382 – 1397.

Henning, K., Jones, A. R., & Holdford, R. (2005). "I didn't do it, but if I did I had a good reason": Minimization, denial, and attributions of blame among male and female domestic violence offenders. *Journal of Family Violence, 20 (3),* 131-139.

Holtfreter, K. & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice,23 (4),* 363 – 382.

Holtfreter, K., Reisig, M. D., & Morash, M. (2004) Poverty, state capital, and recidivism among women offenders. *Criminology and Public Policy, 3(2),* 185-208.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: Wiley-Interscience Publication.

Juhnke, G.A., & Sullivan, T.J. (1995). Attitude changes in DWI offenders: A study of a short-term treatment program. *Journal of Addictions & Offender Counseling, 15(2),* 51-58.

Kaplan, S. J., Pelcovitz, D., & Labruna, V. (1999). Child and adolescent abuse and neglect research: A review of the past 10 years. Part I: Physical and emotional abuse and neglect. *Journal of the American Academy of Child & Adolescent Psychiatry, 38,* 1214–1222.

Keaveny, M.E., & Zauszniewski, J.A. (1999). Life events and psychological well-being in women sentenced to prison. *Issues in Mental Health Nursing, 20,* 73-89.

Kennedy, S. M. (2001). Treatment Responsivity: Reducing recidivism by enhancing treatment effectiveness. In L. L. Motiuk &R. C. Serin (Eds.), *Compendium 2000 on effective correctional treatment* (pp.30-40). Ottawa: Ontario: Research Branch, Correctional Service Canada.

Koons-Witt, B.A., & Schram, P. J. (2003). The prevalence and nature of violent offending by females. *Journal of Criminal Justice, 31,* 361-371.

Kratcoski, P.C., & Scheijerman, K. (1974). Incarcerated male and female offenders' perceptions of their experiences in the criminal justice system. *Journal of Criminal Justice, 2,* 73-78.

Kroner, D. G., & Loza, W. (2001). Evidence for the efficacy of self-report in predicting violent and nonviolent criminal recidivism. *Journal of Interpersonal Violence, 16,* 168-177.

Kroner, D. G. & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28,* 471-489.

Lapham, S.C., Skipper, B.J., Hunt, W.C., & Chang, I. (2000). Do risk factors for re-arrest differ for female and male drunk-driving offenders? *Alcoholism: Clinical and Experimental Research, 24 (11), 1647-1655.*

Lewis, K., Wong, S., & Gordon, A. (2004). *The Violence Risk Scale: Youth Version (VRS-YV).* Department of Psychology and Research, Regional Psychiatric Centre, Saskatoon, SK., Canada: Solicitor General of Canada.

Loeber, R. & Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence: Some common misconceptions and controversies. *American Psychologist, 53,* 242-259.

Loucks, A. & Zamble, E. (1994). Some comparisons of female and male serious offenders. *Forum on Corrections Research, 6 (1),* 22-25.

Loucks, A. D., & Zamble, E. (2000). Predictors of criminal behavior and prison misconduct in serious female offenders. *Empirical and Applied Criminal Justice Review, 1*(1), 1-47.

Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2001). Risk/need assessment, offender classification, and the role of childhood abuse. *Criminal Justice and Behavior, 28,* 543–563.

Loza, W., Neo, L. H., Shahinfar, A., & Loza-Fanous, A. (2005). Cross-validation of the Self-Appraisal Questionnaire: A tool for assessing violent and nonviolent recidivism with female offenders. *International Journal of Offender Therapy and Comparative Criminology, 49,* 547-560.

Luberto, S., Zavatti, P., & Gualandri, G. (1997). Mental illness and criminality: A study of a sample of psychiatric out-patients. In S. Redondo (Ed.), *Advances in psychology and law* (pp.301-311). Berlin: NY: Walter de Gruyter.

Luke, D. A., & Homan, S. M. (1998). Time and change: Using survival analysis in clinical assessment and treatment evaluation. *Psychological Assessment, 10,* 360-378.

Manchak, S. M., Skeem, J. L., Douglas, K. S., & Siranosian, M. (2009). Does gender moderate the predictive utility of the Level of Service Inventory – Revised (LSI-R) for serious violent offenders? *Criminal Justice and Behavior, 36 (5),* 425 – 442.

McGraw, K.O., & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods,1,* 30-46. (Correction 1 (4), 390).

McGee, R. A., Wolfe, D. A., & Wilson, S. K. (1997). Multiple maltreatment experiences and adolescent behavior problems: Adolescents' perspectives. *Development and Psychopathology, 9,* 131–149.

McCartan, L. M. & Gunnison, E. (2010). Individual and relationship factors that differentiate female offenders with and without a sexual abuse history. *Journal of Interpersonal Violence, 25(8),* 1449 – 1469.

McClellan, D. S., Farabee, D., & Crouch, B. M. (1997). Early victimization, drug use, and criminality: A comparison of male and female prisoners. *Criminal Justice and Behavior, 24,* 455 - 476.

Meehl, P. E. (1954). *Clinical versus statistical prediction.* Minneapolis: University of Minnesota Press.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., & Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128–165.

Mills, J. F. & Kroner, D. G. (2006). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. *Criminal Behaviour and Mental Health, 16,* 155- 166.

Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos, 100,* 403 – 405.

Morash, M. (2009). A great debate over using the Level of Service Inventory – Revised (LSI-R) with women offenders. *Criminology and Public Policy, 8(1),* 173-181.

Morris, A., & Wilkinson, C. (1995). Responding to female prisoners' needs. *Prison Journal, 75 (3),* 295-305.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62,* 783 – 792.

Mullen, P. E., Martin, J. L., Anderson, J. C., Romans, S. E., & Herbison, G. P. (1996). The long-term impact of the physical, emotional, and sexual abuse of children: A community study. *Child Abuse & Neglect, 20,* 7–21.

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications (*6th ed.). Upper Saddle River, NJ, USA: Pearson Prentice-Hall, Inc.

Nafekh, M., & Motiuk, L.L. (2002). *The Statistical Information on Recidivism - Revised 1 (SIR-R1) Scale: A Psychometric Evaluation* (Research Report No. R-126). Ottawa: Correctional Service of Canada.

Nichols, D.P. (1998). Choosing an intraclass correlation coefficient. *SPSS Keywords, 67.* Retrieved October 30, 2009, from http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/articles/whichicc.htm

Nuffield, J. (1982). Parole decision making in Canada. Ottawa: Solicitor General of Canada.

O'Keefe, D. J. (2003). Colloquy: Should familywise alpha be adjusted? Against Familywise alpha adjustment. *Human Communication Research,29 (3),* 431 – 447.

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36 (4),* 329 – 353.

Pelissier, B., & Jones, N. (2006). Differences in motivation, coping style, and self-efficacy among incarcerated male and female drug users. *Journal of Substance Abuse Treatment, 30,* 113-120.

Perneger, T. V. (1998). What's wrong with bonferroni adjustments. *British Medical Journal, 316,* 1236 - 1238.

Polaschek, D. L. L., Bell, R. K., Calvert, S. W., & Takarangi, M. K. (2010). Cognitive-behavioural rehabilitation of high risk violent offenders: Investigating treatment change with explicit and implicit measures of cognition. *Applied Cognitive Psychology, 24,* 437 – 449.

Public Safety and Emergency Preparedness Division. (2005). *Sentence calculation: A handbook for judges, lawyers and correctional officials* (3rd ed.*).* Ottawa, ON: Public Works and Government Services Canada.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly, 23(3),* 384 – 405.

Rettinger, L. J. (1998). A recidivism follow-up investigating risk and need within a sample of provincially sentenced women. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.

Rettinger, L. J. & Andrews, D. A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior, 37,* 29 – 46.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63,* 737-748.

Richards, H. J., Casey, J. O., & Lucente, S. W. (2003). Psychopathy and treatment response in incarcerated female substance abusers. *Criminal Justice and Behavior, 30,* 251-276.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology, 1(1),* 43 - 46.

Saskatchewan Corrections & Public Safety. (n.d.). Conditional sentence: Introductory sentence calculation handbook for new sentence administrators. Regina, SK: Adult Corrections Division.

Saskatchewan Corrections & Public Safety. (2004). Introduction to provincial jurisdiction adult sentence calculation: A handbook for criminal justice professionals. Regina, SK: Adult Corrections Division.

Salekin, R. T., Rogers, R., & Sewell, K. W. (1997). Construct validity of psychopathy in a female offender sample: a multitrait-multimethod evaluation. *Journal of Abnormal Psychology, 106,* 576-585.

Salekin, R. T., Rogers, R. Ustad, K. L., & Sewell, K. W. (1998). Psychopathy and recidivism among female inmates. *Law and Human Behavior, 22,* 109-128.

Salisbury, E. J. & Van Voorhis, P. (2009). Gendered Pathways: A qualitative investigation of women probationers' paths to incarceration. *Criminal Justice and Behavior, 36,* 541 – 566.

Salisbury, E. J., Van Voorhis, P., & Spiropoulos, G. V. (2009). The predictive validity of a gender-responsive needs assessment. *Crime and Delinquency, 55 (4),* 550 – 585.

Sandhu, H.S., Dodder, R.A., & Davis, S.P. (1990). Community adjustment of offenders supervised under residential vs. non-residential programs. *Journal of Offender Rehabilitation, 16(1/2),* 139-162.

Serin, R. C., Motiuk, L., & Wichmann, C. (2002). An examination of suicide attempts among inmates. *Forum on Corrections Research 14(2),* 40-42.

Shaw, M. (1994). *Ontario women in conflict with the law: A survey of women in institutions and under community supervision in Ontario*. North Bay: Ontario Ministry of the Solicitor General and Correctional Services.

Shaw, M. & Dubois, S. (1995). *Understanding violence by women: A review of the literature.* Ottawa, ON: Correctional Service Canada.

Sheldon, K. & Krishnan, G. (2009). The clinical and risk characteristics of patients admitted to a secure hospital-based Dangerous and Severe Personality Disorder unit. *British Journal of Forensic Practice, 11(3),* 19 – 27.

Shrout, P.E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Simourd, L. & Andrews, D. A. (1994). Correlates of delinquency: A look at gender differences. *Forum on Corrections Research,* 6 (1), 26-31.

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology, 8(1),* 183 – 208.

Spertus, I. L., Yehuda, R., Wong, C. M., Halligan, S., & Seremetis, S. V. (2003). Childhood emotional abuse and neglect as predictors of psychological and physical symptoms in women presenting to a primary care practice. *Child Abuse & Neglect, 27,* 1247–1258.

Stasiuk, E., Winter, K., & Nixon, B. (1996). *Awaiting trial: Accused persons remanded to custody*. Toronto, Canada: Ontario Ministry of Correctional Services.

Statistics Canada (March 2006). *Uniform Crime Reporting Incident-Based Survey: Reporting Manual.* Ottawa, ON: Canadian Centre for Justice Statistics Policing Services Program.

Statistics Canada (2008). Female offenders in Canada. *Juristat, 28(1),* 1-23.

Stockdale, K. C. (2008). The validity and reliability of the Violence Risk Scale – Youth Version (VRS-YV). Unpublished doctoral dissertation, University of Saskatchewan, SK., Canada.

Strauss, S.M., & Falkin, G.P. (2001). Social support systems of women offenders who sue drugs: A focus on the mother-daughter relationship. *American Journal of Drug and Alcohol Abuse, 27(1),* 65-89.

Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA., US: Allyn and Bacon.

Task Force on Federally Sentenced Women. (1990). *Creating choices: The report of the task force on federally sentenced women.* Ottawa, ON: Task Force on Federally Sentenced Women.

Taylor, K. N., & Blanchette, K. (2009). The women are not wrong: It's the approach that is debatable. *Criminology and Public Policy, 8(1),* 221 – 229.

Topitzes, J., Mersky, J. P., & Reynolds, A. J. (2011). Child maltreatment and offending behavior: Gender-specific effects and pathways. *Criminal Justice and Behavior, 38,* 492 - 510.

Van Voorhis, P., Salisbury, E., Wright, E., & Bauman, A. (2008). *Achieving accurate pictures of risk and identifying gender responsive needs: Two new assessments for women offenders.* Washington, DC: USDOJ National Institute of Corrections.

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior, 37,* 261 – 288.

Verona, E., & Carbonell, J. L. (2000). Female violence and personality: Evidence for a pattern of overcontrolled hostility among one-time violent female offenders. *Criminal Justice and Behavior, 27(2),* 176-195.

Vieira, T. A., Skilling, T. A., & Peterson-Badali, M. (2009). Matching court-ordered services with treatment needs: Predicting treatment success with young offenders. *Criminal Justice and Behavior, 36(4),* 385 – 401.

Vissing, Y. M., Straus, M. A., Gelles, R. J., Harrop, J. W. (1991), Verbal aggression by parents and psychosocial problems of children. *Child Abuse & Neglect, 15,* 223 – 238.

Vitale, J. E. & Newman, J. P. (2001). Using the psychopathy checklist-revised with female samples: Reliability, validity, an implications for clinical utility. *Clinical Psychology: Science and Practice, 8,* 117-132.

Vitale, J.E., Smith, S.S., Brinkley, C.A., Newman, J.P. (2002). The reliability and validity of the Psychopathy Checklist-Revised in a sample of female offenders. *Criminal Justice and Behavior, 29(2),* 202-231.

Vitopoulos, A. N. (2011). The efficacy of the Risk-Need-Responsivity framework in guiding treatment for female young offenders. Unpublished doctoral dissertation, University of Toronto, ON., Canada.

Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes: A meta-analysis. *Criminal Justice and Behavior, 33,* 279 - 304.

Warren, J.I., & South, S.C. (2006). Comparing the constructs of antisocial personality disorder and psychopathy in a sample of incarcerated women. *Behavioral Sciences and the Law, 24,* 1-20.

Weizmann-Henelius, G., Viemero, V., & Eronen, M. (2004). Psychopathy in violent female offenders in Finland. *Psychopathology,37,* 213-221.

Widom, C. S. (1978). An empirical classification of female offenders. *Criminal Justice and Behavior, 5,* 35-52.

Widom, C. S., Schuck, A. M., & White, H. R. (2006). An examination of pathways from childhood victimization to violence: The role of early aggression and problematic alcohol use. *Violence and Victims, 21 (6),* 675 – 690.

Wong, S. (1988). Is Hare's psychopathy checklist reliable without the interview? *Psychological Reports, 62,* 931-934.

Wong, S. C. P., & Gordon, A. (1998 - 2003). *Violence Risk Scale.* (Available from the authors, Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5A5)

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment instrument. *Psychology, Public Policy, and Law, 12,* 279-309.

Wong, S. C. P., Gordon, A., Gu, D (2007). Assessment and treatment of violence-prone forensic clients: an integrated approach. *British Journal of Psychiatry, 190 (suppl.49),* s66 – s74.

Wormith, J. S. (1997). Research to practice: Applying risk/needs assessment to offender classification. *Forum, 9,* 1-8.

Wormith, J. S., & Goldstone, C. S. (1984). The clinical and statistical prediction of recidivism. *Criminal Justice and Behavior,11,* 3-34.

Wright, E. M., Salisbury, E. J., & VanVoorhis, P. (2007). Predicting the prison misconducts of women offenders: The importance of gender-responsive needs. *Journal of Contemporary Criminal Justice, 23,* 310 – 340.

Wright, M. O. (Ed.). (2007). Childhood emotional abuse: Mediating and moderating processes affecting long-term impact [Special issue]. *Journal of Emotional Abuse, 7*(2).

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136 (5),* 740 – 767.